

UNIVERSIDAD AUTÓNOMA DE MADRID
FACULTAD DE FILOSOFÍA Y LETRAS
Programa de Doctorado en Filosofía y Lingüística



LA PROBLEMÁTICA AUTORÍA DE CARLOS CASTANEDA
Un estudio estilométrico de “*candidato único*”

Tesis Doctoral escrita por

Juan Jesús Fernández Trillo

Dirigida por la Doctora Doña Elena Garayzábal Heinze

Madrid, 2015

A mi esposa Sagrario, por su apoyo siempre incondicional y entusiasta.

Agradecimientos

Quiero dar las gracias a la Doctora Elena Garayzábal Heinze por su atenta supervisión y constante estímulo. Sus aportaciones y sugerencias resultaron de enorme ayuda a la hora de resolver las dificultades que a lo largo de la elaboración de esta tesis fueron surgiendo.

ÍNDICE

1. Introducción	1
2. Fundamentos teóricos y herramientas de análisis	11
2.1. El contexto científico de la Lingüística Forense	13
2.1.1. La lingüística y su contribución a la ciencia forense	15
2.1.2. Lingüística forense: Propuesta para una nueva clasificación	19
2.1.2.1. El lenguaje forense	22
2.1.2.2. La atribución lingüística	28
2.1.3. Variación lingüística e idiolecto. El soporte teórico de una hipótesis de trabajo	35
2.1.4. Paradojas y aporías	42
2.2. Estilometría: la estadística al servicio de la lingüística	46
2.2.1. Estilometría y Estilística Forense: Dos aproximaciones a los problemas de estilo e idiolecto	46
2.2.2. Métodos de análisis	56
2.2.2.1. La extensión del texto	58
2.2.2.2. El número de autores	60
2.2.2.3. Los marcadores de estilo	73
A. Caracteres	75
B. Palabras	77
C. Sintaxis	82
D. Frases	84
2.3. La estilometría en acción: Tres casos prácticos	89
2.3.1. <i>The Book of Mormon</i>	90
2.3.2. <i>The Federalist Papers</i>	96
2.3.3. <i>The Cuckoo's Calling</i>	100
3. Carlos Castaneda: Presentación del autor y de su obra	107
3.1. El relato de un aprendizaje	109
3.2. Una biografía llena de sombras	118
3.3. Los elementos de una impostura	126
3.4. Aproximación estilística a la obra de Carlos Castaneda	147

4. Estudio de las obras atribuidas a Carlos Castaneda: Aspectos preliminares	173
4.1. Marco teórico	175
4.2. Antecedentes metodológicos	177
4.3. Determinación del método de análisis empleado	179
4.4. Marco empírico	181
4.4.1. Objetivos del estudio	181
4.4.2. Fases del estudio	182
4.4.2.1. Fase 1: Selección de los marcadores de estilo	183
4.4.2.2. Fase 2: Descripción de las herramientas informáticas	189
4.4.2.3. Fase 3: Pruebas con el Corpus de Entrenamiento	193
4.5. Fase 4: Análisis de las obras indubitadas de Carlos Castaneda.	
Corpus A	203
4.6. Fase 4: Análisis de las obras dubitadas de Carlos Castaneda. Corpus B	208
4.6.1. Análisis de la Obra 1 (Corpus B)	210
4.6.2. Análisis de la Obra 2 (Corpus B)	219
4.6.3. Análisis de la Obra 3 (Corpus B)	228
4.6.4. Análisis de la Obra 4 (Corpus B)	237
4.6.5. Análisis de la Obra 5 (Corpus B)	246
4.6.6. Análisis de la Obra 6 (Corpus B)	255
4.6.7. Análisis de la Obra 7 (Corpus B)	264
4.6.8. Análisis de la Obra 8 (Corpus B)	273
4.7. Fase 5: Determinación de porcentajes de autoría de las obras del Corpus B	282
5. Análisis de los resultados y conclusiones generales	283
5.1. Análisis de los resultados: Sobre la autoría de Carlos Castaneda	285
5.2. Conclusiones generales: Dos factores a considerar en los estudios de atribución de autoría	298
6. Bibliografía	301
6.1. Fuentes primarias consultadas	303
6.2. Fuentes secundarias consultadas	304

1. Introducción

1. **Introducción**

La relación entre identidad y producción lingüística sigue siendo, a pesar de los estudios realizados, un terreno inexplorado en gran medida. Sin embargo, los vínculos que unen la personalidad del individuo con la forma en que este habla o escribe son en muchos casos evidentes, aun cuando se trate de una evidencia obtenida con frecuencia de forma intuitiva.

En el terreno práctico, la constatación precisa y objetiva de dichos vínculos suele resultar una tarea compleja y ello debido a la naturaleza cambiante de los rasgos característicos propios de cada hablante o escritor. Nadie genera discursos en los que los elementos lingüísticos se mantengan constantes a través del tiempo y de las circunstancias que rodean al emisor. La naturaleza misma del lenguaje y su inagotable capacidad para combinar estos elementos mediante recurrencias léxicas, semánticas y sintácticas facilita e impulsa esta diversidad intra-autorial.

El resultado se materializa en una paradoja que combina, por un lado la variación lingüística ejercida por el hablante / escritor y, por otro, una producción textual en la que aparecen regularmente una serie de componentes que permiten individualizarlo de forma eficaz.

A pesar de todo, es preciso reconocer que esta individualización resulta más fácil si el emisor es un hablante, que si es un escritor. Los rasgos fonéticos y fonológicos constituyen en sí mismos una huella distintiva de identidad. Por otro lado, en la comunicación oral interviene también, y de

forma acusada, un uso del lenguaje dictado en gran medida por los hábitos y por la personalidad del hablante. En la escritura, en cambio, se pone en marcha un proceso reflexivo y una selección consciente de los recursos lingüísticos que van a ser empleados en la transmisión del mensaje. Este proceso es aún más intenso, si cabe, cuando se aborda la creación literaria.

El escritor literario escoge cuidadosamente el léxico y la sintaxis, las figuras retóricas y los elementos ortográficos. Todo ello hace que el resultado corresponda menos a su idiosincrasia psicológica y más a un esfuerzo consciente por obtener un efecto concreto en el lector. Pero incluso en estos casos, existen estilos asociados a determinados escritores, que el público reconoce y demanda. A pesar de no poder determinar de forma precisa, en muchos casos, qué es lo que hace diferentes a los textos de un escritor concreto, el lector, de nuevo de forma intuitiva, identifica a su autor favorito e incluso, tras algún tiempo leyendo sus obras, es capaz de asegurar si una pieza literaria que lleva su nombre ha sido elaborada por algún otro.

Esto no siempre es así, por supuesto; no todos los escritores poseen un estilo claramente reconocible, ni todos los lectores disponen de una percepción literaria capaz de identificar ni tan siquiera aquellos autores más idiosincrásicos. Y es precisamente en este punto donde interviene la lingüística forense, ciencia relativamente reciente que pretende, entre otros objetivos, determinar la autoría de un texto a partir de su análisis estilométrico.

La estilometría, por su parte, descompone por medio de mediciones estadísticas un texto escrito y pone los resultados a disposición del lingüista, quien debe encontrar, en las repeticiones de ciertos elementos, patrones regulares característicos de un autor particular. Se trata de un método que por su propia esencia proporciona resultados mensurables, reproducibles y, en consecuencia, objetivos.

La atribución de autoría basada en métodos estilométricos, sin embargo y al igual que el resto de la lingüística forense, tiene por el momento escaso recorrido en el tiempo. Ciertas cuestiones todavía no han obtenido respuesta satisfactoria, a pesar de que cada día nuevos estudios, métodos de análisis y herramientas técnicas contribuyen a consolidar este campo del conocimiento. Una de estas cuestiones, tal vez la más importante, es el alcance de los análisis de atribución de autoría; esto es, hasta dónde se puede llegar en el esfuerzo por atribuir un texto a un escritor determinado. En este aspecto, confiamos que los resultados del presente trabajo de investigación constituyan una aportación valiosa.

Para ello se ha tomado como materia de estudio la obra del escritor Carlos Castaneda y se ha llevado a cabo un análisis estilométrico exhaustivo de las doce obras publicadas bajo su nombre. Ciertas diferencias formales y de contenido dividen el conjunto de su obra en dos partes a las que se ha denominado a efectos prácticos Corpus A y Corpus B, el primero de ellos integrado por los cuatro primeros libros y el segundo por los ocho posteriores. En estos últimos, en el Corpus B, se encuentran diferencias significativas con respecto a los primeros, que constituyen el Corpus A, diferencias que podrían explicarse mediante la intervención de un segundo autor.

El objetivo de esta tesis, por tanto, consiste en determinar si existen elementos objetivos que permitan afirmar con certeza que alguna de las obras que integran el denominado Corpus B no fue escrita por el propio Carlos Castaneda.

La ejecución práctica de este análisis reúne, además, dos aspectos que a la vez facilitan y dificultan la tarea. Por un lado, se dispone de una gran cantidad de texto, tanto en las obras dubitadas como en las indubitadas. Esto no es frecuente y por supuesto posibilita la extracción de elementos característicos tanto del escritor analizado, como de los textos cuya atribución

se intenta verificar. Uno de los principales obstáculos en este tipo de tareas es precisamente que el texto no sea lo suficientemente extenso como para definir con claridad el estilo de su autor. El otro aspecto al que nos referimos, en cambio, complica la tarea, pues al existir un solo autor candidato resulta ineficaz el método habitual de atribuir el texto dubitado al autor del texto que más se le parece.

Así pues, este es el caso con que nos encontramos aquí: un único autor, ciertas obras literarias y la tarea de determinar si éstas han sido escritas o no por aquél.

Como ya hemos dicho, el autor en cuestión es Carlos Castaneda. En España, los libros de Castaneda comenzaron a publicarse en la década de los setenta, algunos años después de que lo hiciesen en Estados Unidos, y al igual que sucedió allí – y en el resto de los países donde fueron editados – se convirtieron en una especie de *best-sellers* alternativos. Eran los años de la contracultura¹, que mezclaba aspectos aparentemente tan dispares como el consumo de drogas psicodélicas con el misticismo oriental. En ese ambiente propicio, los libros de Castaneda recibieron una excelente acogida. De acuerdo a su propio testimonio, el escritor, antropólogo de la Universidad de California, UCLA, había entrado en contacto y se había convertido en discípulo de un anciano indio yaqui, heredero de una antigua tradición y poseedor de una serie de conocimientos que le permitían acceder a una realidad alternativa, en parte gracias al consumo de ciertas plantas alucinógenas y en parte por medio de la manipulación de los elementos que encontraba en sus excursiones didácticas por los desiertos de Sonora, donde tenía lugar el aprendizaje.

¹ El término “contracultura” fue acuñado por el académico y profesor de historia en la Universidad de California, Theodore Roszak, con la aparición de su libro *The Making of a Counter Culture* (1969), apenas un año más tarde de que Carlos Castaneda publicara su primera obra.

Narrados en primera persona y estructurados en la forma de un diario de campo donde el investigador iba dejando constancia de sus conversaciones con su maestro y de las indicaciones que este le daba, los libros de Castaneda contenían algo más que los detalles de su enseñanza; cada nuevo ejemplar contribuía a establecer un sistema de creencias, complejo pero coherente, del que nadie había oído nunca hablar. Este sistema daba respuestas a las principales preguntas existenciales y proporcionaba, a la vez, normas de comportamiento para hacer frente a los avatares de la vida. Por otro lado, la particular técnica narrativa de Castaneda y su diestra manipulación de los elementos de la historia, convertían a esta en un relato emotivo, que rozaba en ocasiones lo poético, lo cual contribuyó, sin duda y en buena medida, a su éxito entre los lectores de una generación determinada.

Este trabajo se organiza de acuerdo a una división en distintas secciones ordenadas con la finalidad de que cada una prepare al lector para las que siguen. Así, tras la presentación anterior y esta misma descripción de la estructura de la tesis, se da paso a una revisión de los fundamentos de la lingüística forense y de sus herramientas más importantes. Se pone en contexto el alcance de su contribución a la ciencia forense en general, destacando el hecho de que, a pesar de su reciente incorporación a los procesos judiciales, tanto los análisis lingüísticos de los textos legales como los estudios de atribución de autoría han venido ganando relevancia en los últimos años.

Dado que la atribución de autoría centra el objeto de esta tesis, se le presta una atención especial comparándola con técnicas similares aplicadas en otras áreas forenses, como la dactiloscopia o el análisis del ADN. Nos ha parecido importante incidir, además, sobre el valor real de dichas pruebas y el grado de fiabilidad que todas ellas poseen, resaltando que en ningún caso y por elevado que sea alcanzan una precisión absoluta.

Se revisa, a continuación, el estado de la lingüística forense en nuestro país destacando la tarea de algunos investigadores que, además de abrir el camino a esta joven disciplina, la han hecho avanzar incluso en el terreno internacional.

Se propone una nueva clasificación de los campos o especialidades abarcadas por la lingüística forense, que podría sustituir a la clasificación que viene siendo empleada actualmente y se centra la atención en la estilometría, o el análisis estadístico de textos, herramienta de gran valor en la tarea de atribución de autoría.

A continuación se examinan los presupuestos teóricos que soportan las investigaciones en este campo: la variación lingüística y el idiolecto. Tanto uno como otro – de hecho el segundo es una consecuencia del primero – han sido muy estudiados a lo largo del pasado siglo y existe en la actualidad una base suficiente, tanto empírica como teórica, para afirmar que cada individuo posee una forma idiosincrásica de expresarse. Este hecho es el punto de partida del lingüista forense cuando aborda la tarea de atribuir un texto a un autor concreto, sea mediante análisis estilísticos o mediante estudios estilométricos.

En este punto se reflexiona sobre las implicaciones y alcance de los supuestos mencionados anteriormente. En ocasiones, se ha comparado la “huella lingüística” con la huella digital o con el ADN, en un intento de poner de manifiesto las posibilidades que ofrecen este tipo de estudios, generando tal vez unas expectativas que no se ajustan del todo a la realidad.

Se describen a continuación y en detalle las técnicas estilométricas dado que serán las empleadas en el presente trabajo, toda vez que se repasa la historia reciente y los estudios más señalados en el campo de la estilometría y atribución de autoría y que han servido de punto de partida a investigaciones

posteriores. Por otro lado, una rápida revisión al estado de la cuestión pondrá de manifiesto la diversidad de métodos y sistemas aplicados, dentro de este campo. Esta diversidad, si bien ofrece un gran abanico de posibilidades al investigador, presenta la contrapartida de lastrar su avance, generando en ocasiones controversias que obstaculizan su pleno reconocimiento como elemento probatorio en el proceso judicial.

Algunos de los factores implicados en todo análisis estilométrico reciben una especial atención, debido a su importancia. La extensión del texto que se analice o los marcadores elegidos son aspectos a tener muy en cuenta a la hora de abordar esta tarea, como lo es el número de autores candidatos. Dada la naturaleza de los análisis estilométricos, resulta crucial saber de antemano con certeza si el autor del texto cuestionado – o dubitado – se haya o no entre los autores candidatos o “sospechosos”. De ello, y de que su número sea lo más reducido posible, dependerá en gran medida la precisión del juicio emitido finalmente.

Por otro lado, los estudios de atribución de autoría han sido en los últimos tiempos abordados, en la mayor parte de los casos, por expertos en ciencias matemáticas, estadística o programación informática, relegando hasta cierto punto a los lingüistas a un segundo plano. La complejidad de los procesos empleados ha dificultado la supervisión de algunos de dichos estudios. Sin embargo, el convencimiento de que es posible llevar a cabo análisis estilométricos que empleen técnicas estadísticas sencillas y comprensibles, incluso para quienes carecen de conocimientos especializados, nos ha llevado a ejemplificarlo en la práctica con el presente trabajo. Para lograrlo nos hemos apoyado en un razonamiento elemental: la demostración de que dos elementos, diferentes en apariencia, han tenido un mismo origen requiere encontrar en ambos elementos comunes. Este es el principio que guía este estudio contrastivo.

Es cierto que un creador puede elaborar dos piezas – sean textos o de cualquier otra naturaleza – sin ningún rasgo en común entre ellas, pero en ese caso, tomadas de forma aislada y sin conocer quién fue su artífice, será imposible determinar si son obra o no del mismo autor. En otras palabras, no debe ser imposible para un escritor elaborar dos textos por completo distintos entre sí gracias a un esfuerzo consciente, pero si lo hace, quien los analice será incapaz de determinar si fueron escritos por la misma persona.

En esta tesis se mostrarán estudios de frecuencia y valores medios de aparición de ciertos rasgos estilísticos, en la mayoría de los casos por medio de gráficos que permitirán apreciar visualmente similitudes y diferencias. Todo ello convierte tanto la fase de análisis como la de obtención de conclusiones en un proceso asequible para cualquiera que aplique razonamientos lógicos elementales.

En conjunto confiamos en que el resultado y la minuciosa descripción del proceso seguido serán de utilidad, no solo para nosotros mismos, sino también para todo aquel que desee continuar investigando en el terreno de la atribución de autoría, el idiolecto literario y la relación entre identidad y producción lingüística.

2. Fundamentos teóricos y herramientas de análisis

2.1. El contexto científico de la Lingüística Forense

En la sección anterior formulábamos el objetivo principal que nos hemos fijado al comenzar este trabajo, esto es, verificar o descartar la hipótesis según la cual todas o algunas de las ocho obras firmadas por Carlos Castaneda, que integran lo que hemos denominado Corpus B, fueron escritas por un autor diferente del que escribió las cuatro primeras, a las que hemos denominado Corpus A. Esta tarea, la atribución de autoría literaria de un texto o conjunto de textos sobre los que existen dudas, si bien no es infrecuente entre lingüistas y críticos literarios, siempre resulta controvertida debido sobre todo a cuestiones de método, que, a su vez, dan lugar a resultados cuyas conclusiones son, en ocasiones, discutibles.

No es de extrañar, por tanto, que la investigación académica en este campo aparezca fragmentada. A ello contribuyen, además, tanto la diversidad de objetivos perseguidos (identificación de autoría de notas de suicidio, amenaza o extorsión (Abbasi y Chen, 2005); de autoría de textos literarios en sus distintos géneros (Burrows, 2002; Hoover, 2004); disputas civiles sobre derechos de marcas (Chaski, 2005, Grant, 2007); de autoría de textos extraídos de internet: blogs, mails, posts, etc. (Luyckx, 2010) como la variedad e irregular eficacia de los diferentes métodos propuestos (Koppel, Schler, Argamon y Messeri, 2006), el número de autores candidatos o los distintos marcadores de estilo utilizados en cada estudio (Grieve, 2007). Por ello, creemos conveniente un recorrido por los principales métodos de análisis y las teorías que los soportan, para hacernos siquiera una idea de las posibilidades que cada uno ofrece.

En las próximas páginas, pues, expondremos los fundamentos de la ciencia forense y del lugar que en ella ocupa la lingüística forense, de los elementos que constituyen esta nueva disciplina y de las bases teóricas que la sustentan. Expondremos cómo los planteamientos de los que parten algunas de las más importantes técnicas de identificación, presentan muchas similitudes con las técnicas empleadas para la atribución de autoría.

Hablaremos, aunque solo sea de forma somera, de la naturaleza de la Estilometría, disciplina reciente dentro de la Lingüística Forense, pero que ha realizado ya esfuerzos considerables por sistematizar sus métodos de investigación y por ofrecer resultados objetivos y verificables, algo necesario, por lo demás, cuando pretende que dichos resultados sean empleados como evidencias periciales dentro del proceso judicial.

Resultados objetivos y un proceso de obtención susceptible de ser replicado son, pues, dos máximas que acompañan a la Lingüística Forense en todo momento. Por otro lado, el paradigma cuantitativo que guía el método estilométrico genera datos que son analizados estadísticamente, lo que proporciona consistencia y fiabilidad a las conclusiones finales.

Así pues, describiremos las principales herramientas informáticas que se emplean habitualmente en los análisis estilísticos y con los que se busca determinar la autoría de textos dubitados. Mostraremos, además, el grado de fiabilidad que ofrecen cada una de estas herramientas, fiabilidad que varía dependiendo de factores como la extensión de los textos analizados, el tipo de texto de que se trate y del número de posibles autores. Expondremos igualmente las discrepancias entre los diferentes especialistas en Lingüística Forense acerca de la conveniencia del empleo de uno u otro, pues, al igual que sucede en otros campos de la ciencia del lenguaje, no existe por el momento un criterio unificado a este respecto.

A continuación expondremos el método de análisis elegido y las razones para esta elección, condicionado por las especiales circunstancias del estudio que nos ocupa, esto es, textos muy extensos y tan solo un autor candidato. El algoritmo que emplearemos es en buena medida innovador y por ello nos ha parecido conveniente describir en detalle cada paso del proceso.

2.1.1. La lingüística y su contribución a la ciencia forense

La ciencia forense busca, de forma general, contribuir al proceso judicial, colaborando en la reconstrucción de los hechos investigados e identificando a quienes tomaron parte en ellos. Se trata de una disciplina transversal que se nutre de áreas muy diversas del conocimiento, desde la medicina hasta la química, pasando por la biología, la antropología o las ciencias del comportamiento y, desde hace algún tiempo, la lingüística. En todos los casos, los peritos buscan rastros, evidencias que proporcionen respuestas a las preguntas que todo investigador se formula: Qué, cuándo, cómo, por qué y, sobre todo, quién.

La ciencia forense se sustenta en el llamado Principio de Transferencia, que establece que cuando dos cuerpos entran en contacto cada uno de ellos deja su impronta en el otro². Esto significa, por ejemplo, que las ruedas de un vehículo dejan marcas de rodadas con el dibujo de los neumáticos en una pista de tierra y, a la vez, parte de esa tierra se adhiere a los neumáticos, lo que demostrará de forma fehaciente que dicho vehículo estuvo, en algún momento, en cierto lugar concreto. En la práctica, sin embargo, el proceso no

² También conocido como Principio de Locard, en recuerdo del doctor Edmond Locard, quien a comienzos del siglo XX, estableció la máxima forense que guiaría a los investigadores futuros, según la cual todo contacto deja un rastro y si no lo vemos es debido a que no hemos mirado adecuadamente (Kirk, 1953). Este principio, junto al concepto filosófico de *causalidad* – todo efecto ha sido producido necesariamente por una causa anterior – ha inspirado a los investigadores policiales a lo largo de la historia y a cientos de escritores del género detectivesco.

siempre resulta tan evidente; la pista por la que el vehículo circula puede ser de duro asfalto seco, por lo que la transferencia de elementos entre neumáticos y pista tiene lugar a un nivel tan infinitesimal que somos incapaces, por el momento, de encontrar evidencias de la misma. Por fortuna, en la práctica tampoco es necesario aportar rastros que demuestren esta transferencia en ambos sentidos. Es suficiente con que uno de los dos cuerpos deje en el otro una marca, una señal perceptible por nuestros sentidos o por nuestros instrumentos. Nos basta, por ejemplo, con encontrar una huella dactilar en la superficie de un vaso para demostrar la identidad de la persona que lo sujetó en un momento dado; no necesitamos encontrar restos del polvo que cubría el vaso en la mano del sospechoso.

Precisamente el estudio de las impresiones dejadas por las huellas dactilares, o dactiloscopia, constituye un capítulo fundamental dentro de la ciencia forense. De entre todas las técnicas de identificación personal, la dactiloscopia ha adquirido la fama merecida de ser la más eficaz, con la excepción de las recientes técnicas de análisis de ADN. Esa disciplina parte de los estudios que Cesare Lombroso y Alphonse Bertillon³ llevaron a cabo durante el siglo XIX, intentando encontrar elementos físicos que permitiesen identificar de forma inequívoca a los delincuentes. Serían el británico Francis Galton (1822-1915) y el argentino Juan Vucetich (1858-1925) quienes establecerían las bases de lo que llegaría durante el siglo XX a convertirse en el método más utilizado para reseñar a los delincuentes y para vincularlos con la escena del crimen.

³ Ezequiel Marco “Cesare” Lombroso (1835-1909) había formulado una teoría que durante parte del siglo XIX tuvo muy buena acogida: la de que las tendencias criminales del ser humano encontraban su reflejo en una fisionomía determinada. Así, Lombroso catalogó los rasgos que, según él, señalaban sin lugar a dudas una “naturaleza criminal innata”. Su método, basado en la observación empírica, recibió el nombre de positivismo criminológico y, a tono con el espíritu naturalista de la época, despojaba al criminal de toda responsabilidad moral, dado que eran sus instintos naturales los que le empujaban a delinquir (Lombroso, 2009). Por su parte Alphonse Bertillon (1853-1914) elaboró un catálogo de medidas físicas, tales como longitud de las extremidades, separación entre los ojos, circunferencia del cráneo, etc. con las cuales individualizaba a los sujetos reseñados (Rhodes, 1956). Su método, la antropometría, basado en la teoría de que no existen dos personas físicamente idénticas, no es muy diferente, en esencia, de la filosofía que anima los estudios de lingüística forense, como veremos.

La investigación criminal, imbuida por el espíritu científico de la época, buscaba una forma de probar objetivamente que un sospechoso era el autor de los hechos de los que se le acusaba. De manera muy similar a los lingüistas forenses de la actualidad, los investigadores precisaban de evidencias irrefutables de autoría y las hallaron en las huellas digitales dejadas por los criminales en la escena del delito. Los estudios de Galton y Vucetich establecían que los dibujos de las crestas papilares que cada ser humano lleva en la punta de los dedos son únicas e irrepetibles. La dactiloscopia establece que las huellas son inmutables, ya que aparecen durante la formación del feto y permanecen hasta el fallecimiento de su propietario, y diversas, pues no hay dos idénticas.

Estas características revisten a las evidencias dactiloscópicas del peso probatorio necesario para ser empleadas como evidencia en un juicio, pero la dactiloscopia fue un poco más allá: estableció un sistema de clasificación basado en una serie de marcas características cuya presencia, o ausencia, en una huella dactilar – “fingerprint” – da origen a una combinación que, transformada en una fórmula, se convierten en una seña de identidad vinculada a un único individuo en el mundo⁴. Este sistema de identificación se ha convertido en un ideal al que los lingüistas forenses aspiran y les ha llevado, en ocasiones, a calificar sus técnicas de “writeprint” o “voiceprint” – “huella escrita” o “huella vocal” – en un intento de equiparar la fiabilidad de los primeros con la de los segundos. En ocasiones han empleado también la expresión “linguistic DNA”, lo que lleva la analogía, tal vez, un poco demasiado lejos⁵.

⁴ Los términos absolutos, sin embargo, deben ser tomados con precaución, en el terreno científico. En realidad los criminólogos parten de la idea de que no *debería* haber dos huellas iguales dado que las probabilidades matemáticas de que así fuera son elevadísimas. Algunos autores, como Simon A. Cole (2002), aconsejan recordar que “altamente improbable”, no equivale a “imposible”, e instan a seguir mejorando los sistemas de identificación personal.

⁵ Coulthard previene contra el uso del término “linguistic fingerprint” y las conclusiones excesivamente optimistas a que puede dar lugar. Como decíamos, la lingüística forense, en el aspecto de atribución de autoría, queda todavía un poco lejos de la dactiloscopia, tanto en logros, como en reconocimiento en ámbitos judiciales (Coulthard, 2005b).

El principio que sigue la genética forense se basa, de nuevo, en la diversidad: No existen dos seres humanos – salvo los gemelos univitelinos – que posean la misma combinación genética. Este polimorfismo – para emplear el término habitual en esta rama científica – permite individualizar las muestras recogidas de la escena donde se ha cometido un crimen (sangre, semen, cabellos, saliva, uñas o, en general, cualquier tejido orgánico) y compararlo con el de los sospechosos con la certeza de que, si el autor se encuentra entre ellos, los análisis lo señalaran sin el más mínimo rastro de duda. Aunque, si deseamos ser más precisos, debemos reconocer que nos encontramos ante una situación muy similar a la que veíamos en el caso de la dactiloscopia: La certeza con la que un análisis de ADN señala al autor de los hechos investigados es, aproximadamente, del 99’73 %. Si esto es suficiente o no para emitir un juicio depende de lo exigentes que decidamos ser, pero en general se considera que un resultado *positivo* en un análisis de ADN es incuestionable (Goodwin et al., 2007).

Más arriba decíamos que equiparar la fiabilidad de los resultados de la lingüística forense con los del análisis del ADN es pecar, tal vez, de benévolos. Los porcentajes de certeza que ofrecen los análisis estilométricos quedan todavía un poco por detrás de los que alcanza el análisis genético. Sin embargo, hay un aspecto en el que esa distancia se reduce considerablemente: a la hora de descartar un sospechoso o candidato. El ADN permite eliminar con una certeza del 100 % a un individuo sospechoso (Buckleton et al., 2005) y de manera muy similar, la lingüística forense es mucho más precisa cuando se trata de descartar la autoría, que cuando tiene que atribuirle.

Ya vemos, pues, la importancia que la ciencia forense otorga a la atribución de autoría y para lograrlo se esfuerza por encontrar evidencias irrefutables que vinculen el *acto* con el *actor*. De igual forma, los lingüistas forenses basan su labor en la convicción de que un proceso muy similar al Principio de Transferencia tiene lugar durante la producción lingüística, ya

que consideran que en todo texto, sea hablado o escrito, es posible encontrar ciertos rasgos que nos permiten identificar a su autor. Se trata de elementos lingüísticos que son empleados de forma más o menos inconsciente y que convenientemente aislados, medidos y contrastados señalan al autor con un elevado grado de probabilidad.

2.1.2. Lingüística forense: Propuesta para una nueva clasificación

En 1950, en South Wales, Inglaterra, Timothy John Evans fue juzgado, declarado culpable y colgado en la horca por las muertes de su esposa e hija. La sentencia se fundaba, principalmente, en las declaraciones autoinculpatorias que el acusado había realizado ante la policía, meses antes del juicio. A pesar de que durante la vista oral el acusado se retractó y señaló a su vecino John Christie como responsable de los homicidios, Evans fue condenado y ejecutado. Tres años después, la policía detuvo John Christie, acusado de ser un asesino en serie. Entre los homicidios que pudieron ser demostrados y que él mismo reconoció estaban los de la esposa e hija de Evans, por lo que a este se le concedió un perdón póstumo. Quince años más tarde, en 1968, el lingüista Jan Starvirk realizó un estudio de los testimonios realizados por Evans tras ser arrestado, en el que demostraba que la “confesión de culpabilidad” que la policía había presentado como prueba principal estaba elaborada a partir de diversas declaraciones realizadas por Evans en distintos momentos (Olsson, 2008: 22). Starvirk publicó sus conclusiones en *The Evans Statements: A Case for Forensic Linguistics*, y su trabajo está considerado como el origen de la moderna lingüística forense⁶ (Coulthard y Johnson, 2007; Garayzábal et al., 2012; McMenamin, 2002; Olsson, 2008).

⁶ Al parecer el término lingüística forense había sido ya utilizado con anterioridad por Philbrick, en 1949, como título de un estudio de la lengua inglesa en el ámbito legal, pero la expresión no arraigó hasta que fue publicado el trabajo de Svartvik, en 1968 (Coulthard y Johnson, 2007).

Relativamente reciente como es, a la lingüística forense todavía le queda algún camino por recorrer hasta alcanzar la sistematización y, especialmente, el reconocimiento en el ámbito jurídico que actualmente poseen el resto de las disciplinas que integran la ciencia forense. Como suele ser habitual en los comienzos, la diversidad de opiniones y puntos de vista entre los especialistas en cuanto a sus límites, materia de estudio y métodos de trabajo difumina ligeramente la imagen de conjunto. A pesar de ello, sí que existe una visión clara de hacia dónde debe enfocar su mirada investigadora y en la definición que de forma genérica proporciona la Asociación Internacional de Lingüistas Forenses (IAFL) se resume el objeto de su estudio: la interfaz entre lengua y Derecho.

A partir de ahí, las definiciones que unos u otros proporcionan varían según la importancia que otorguen a las distintas materias o áreas que la integran. Así, por ejemplo, Sol Azuelos-Atias (2007) se centra en el discurso jurídico, del que realiza un análisis pragmático, mientras que Olsson lo hace en el análisis del lenguaje escrito, subrayando que cualquier texto, desde una nota de rescate a un *ticket* de autopista, puede ser objeto de estudio por parte de la lingüística forense, empleando para ello todos los recursos lingüísticos a disposición del investigador:

“(...) the forensic linguist may quote observations from research undertaken in fields as diverse as language and memory studies, Conversation Analysis, Discourse Analysis, theory of grammar, Cognitive Linguistics, Speech Act Theory, etc.” (Olson, 2008: 4-5).

Por su parte, Gibbons y Turell (2008) sugieren una triple división de las áreas de estudio: los textos legales y la legislación en general, el discurso legal de los juzgados y los interrogatorios policiales y el estudio de las

evidencias lingüísticas. En nuestro país, la Doctora M^a Teresa Turell⁷, desde la Universidad Pompeu-Fabra de Barcelona, desarrolló en los últimos años una muy importante labor en la consolidación de la lingüística forense en España y su trabajo se ha convertido en una referencia obligada para los investigadores en este campo. Turell establece la siguiente clasificación de la lingüística forense:

- a. El lenguaje de la ley
- b. El lenguaje judicial
- c. El lenguaje probatorio o evidencial

Esta división, heredada de los estudios anglosajones, se ha adoptado en nuestro país (Braz, 2011; Turell, 2008, 2011; Santana y Falces, 2012), sin duda por la autoridad de los autores que la respaldan y por su pragmatismo a la hora de estructurar la disciplina en cuestión. Sin embargo, nos parece que podemos contribuir a la sistematización de la lingüística forense proponiendo una clasificación que, si bien introduce algún cambio menor, ordena de una forma más racional las áreas de estudio.

En nuestra opinión, es posible establecer dos grandes campos de trabajo en función de la materia de la que se ocupan: por un lado el uso del lenguaje en el ámbito legal y judicial y por otro la atribución de autoría de textos y grabaciones de voz. Cada uno de estos campos agrupa tareas muy próximas entre sí, que comparten planteamientos y propuestas teóricas. De esta manera la división de la Lingüística Forense quedaría como sigue:

- 1) El lenguaje forense
 - a. Lenguaje legislativo
 - b. Lenguaje judicial

⁷ La Doctora M^a Teresa Turell falleció en el año 2013 y sus alumnos y discípulos continúan con el trabajo de investigación que ella inició, en el ForensicLab, de la Universidad Pompeu-Fabra.

- c. La traducción jurídica
 - d. Lenguaje policial
-
- 2) La atribución lingüística
 - a. Fonética, fonología y acústica forense
 - b. Detección de plagio
 - c. Estilometría

2.1.2.1. El lenguaje forense

El estudio del lenguaje jurídico acumula en el mundo anglosajón una tradición considerable, si se la compara con la de nuestro país. Los primeros trabajos en este campo se remontan a la década de 1930 (McMenamin, 2002), pero en las últimas décadas han proliferado las publicaciones especializadas, de entre las que destacan los estudios bajo la perspectiva del análisis del discurso legal de Roger W. Shuy (1984, 1986, 1993, 1998). En España, se está empezando a tomar conciencia de la importancia que para el correcto funcionamiento de las instituciones democráticas en general y la aplicación de la justicia en particular tiene una redacción adecuada de normas y leyes.

“El análisis del discurso proporciona, entre otras cosas, una visión de la complejidad de la jerga legal o lenguaje jurídico, que puede ayudar a solventar problemas de interpretación de determinados textos dentro de este ámbito. En muchas ocasiones, el uso de elementos retóricos y arcaicos, la naturaleza de los términos utilizados, o simplemente una mala redacción, pueden hacer que una ley sea ininteligible y requiera un proceso de interpretación para su correcta aplicación”. (Garayzábal et al., 2012: 33).

Quienes se ocupan del estudio del **lenguaje legislativo** examinan los textos legales, con la intención de señalar aquellos aspectos de los mismos que pueden ofrecer problemas de interpretación, bien por su ambigüedad, bien por una excesiva complejidad. Una comprensión clara de las leyes, normas y disposiciones que regulan la vida pública y privada se traduce en una sociedad más justa e igualitaria, en la que las diferencias sociales y culturales se atenúan.

En este sentido es conveniente destacar los esfuerzos que en los países de habla inglesa se vienen llevando a cabo desde hace años a través de los Índices de Legibilidad destinados a que no solo leyes, sino normativa de todo tipo, tanto pública como privada, se redacte en términos comprensibles para todo el mundo⁸. Distintos tipos de test, como el *Flesh-Kincaid* (Kincaid et al., 1975), la fórmula *Dale-Chall* (Dale y Chall, 1948), o el índice *Gunning Fog* (Gunning, 1952, 2006) proporcionan sistemas de medida del grado de comprensibilidad de un texto, a partir de la mayor o menor frecuencia de aparición en el lenguaje cotidiano de las palabras empleadas en dicho texto. Factores como la longitud de las palabras o de las frases son también tenidas en cuenta a la hora de proporcionar el mencionado índice de legibilidad⁹. En España, la Comisión para la Modernización del Lenguaje Jurídico (CMLJ) está empezando la que sin duda será una ardua y penosa tarea en este campo, intentado proporcionar recomendaciones que hagan más accesibles los textos jurídicos que regulan la vida de todos los españoles. El estudio de Mercedes

⁸ Es interesante recordar aquí casos recientes en los que una comprensión deficiente de las condiciones de las cláusulas contractuales (hipotecas con “cláusulas suelo” o “participaciones preferentes”- ABC, 10-07-2013; El País, 11-07-2013; Expansión, 23-07-2013; El Mundo, 03-07-2013) ha llevado a muchos ciudadanos, pequeños inversores de escasa formación cultural, a perder buena parte de sus ahorros. La tradicional “letra pequeña” de los contratos de pólizas de seguros, que en ocasiones oculta condiciones desfavorables para los firmantes, es otro buen ejemplo de un acceso difícil –en este caso intencionado – a las condiciones de los acuerdos comerciales.

⁹ Durante el año 2013 la Universidad de Nottingham puso en marcha el programa *Literatin*, de acceso y descarga libre, destinado a determinar el grado de legibilidad de todo tipo de textos. Como ejemplo de aplicación de este programa, examinaron las condiciones legales que aparecen en los productos Google y como resultado concluyeron que tan solo el 43% de la población adulta de Inglaterra podría comprenderlas sin dificultad (Hood, 2013).

Bengoechea (2014) proporciona un completo estudio de la situación en nuestro país, de las posibles causas de la oscuridad que tradicionalmente ha caracterizado a este tipo de textos, así como de las iniciativas y propuestas de la citada CMLJ para avanzar hacia la formulación de leyes cuya comprensión sea accesible a toda la ciudadanía, en general.

El **lenguaje judicial**, por su parte, ofrece problemas muy similares, salvo que en este caso el proceso comunicativo suele desenvolverse en buena parte en el plano oral. El procedimiento jurídico, de por sí ajeno a la experiencia del ciudadano medio¹⁰, acostumbra a estar plagado de tecnicismos y expresiones formuladas por abogados, procuradores, fiscales y magistrados que, en muchos casos, resultan de difícil comprensión. La consecuencia más inmediata es la de situar a los demandantes, imputados y litigantes en general en una posición de indefensión. En pocas ocasiones como durante el proceso judicial se da un desempoderamiento del ciudadano tan acusado. Privado de la palabra, ya que sus manifestaciones y alegaciones se realizan en su mayor parte a través de su abogado, el individuo ve además dificultada su comprensión de lo que está sucediendo debido a la jerga empleada por el resto de los actores. Ciertos sectores sociales, como menores de corta edad, inmigrantes con un conocimiento deficiente de la lengua dominante o personas de escasa formación, por ejemplo, requieren una atención especial para que sean conscientes siquiera de los derechos que les asisten (Atkinson y Drew, 1979; Bennet y Feldman, 1981).

La **traducción jurídica** se vincula de forma directa al lenguaje judicial y cobra una especial relevancia en todas aquellas situaciones en las que “una acusación se fundamenta en hechos lingüísticos que requieren los servicios de un traductor o intérprete” (Garayzábal et al., 2012: 36). Labor siempre

¹⁰ Agravado, además, por el conocimiento tergiversado y del todo punto fuera de contexto que las series de televisión y películas procedentes de los Estados Unidos proporcionan de la legislación y procedimientos judiciales.

delicada, la traducción se convierte en un elemento crucial para aquellos que no dominan el idioma en el que el juicio se desarrolla, que suelen ser, además, los elementos más vulnerables de todos los que intervienen el proceso judicial. De su importancia da idea el hecho de que “la primera norma reguladora de la tarea del traductor jurídico en nuestro país data del año 1529” (Jiménez, 2012: 144). Por desgracia, la situación de quienes desarrollan esta relevante labor es de una ausencia casi total de legislación reguladora o siquiera de unanimidad a la hora de referirse a la misma. En su ensayo “Estudios de lingüística, traducción e interpretación forenses”, Miriam Jiménez propone agrupar tareas y denominaciones:

“Nuestra propuesta es la de reunir bajo la denominación de Traducción Forense e Interpretación Forense todas las posibles tareas que el traductor o el intérprete realizan y en las que existe una relación con el ámbito legal. Estas disciplinas tendrían como objeto de estudio principal, por tanto, la interrelación entre las actividades de traducción e interpretación y el Derecho. Mantendrían un enfoque multi- e interdisciplinar, incluyendo tanto las teorías lingüísticas como los estudios traductológicos sobre el tema y las herramientas que las disciplinas afines (Derecho, Psicolingüística, Sociolingüística, Análisis del Discurso y Pragmática, entre otras) puedan aportar”. (Jiménez, 2012: 47-48).

En la traducción jurídica, o forense para ajustarnos a la propuesta de Jiménez, intervienen conocimientos que van más allá de los puramente lingüísticos, pues la diversidad semántica no es sino un reflejo de la distancia cultural de quien habla otro idioma, factor que sin duda influye en la interpretación de lo que en una vista oral se dice. El discurso legal y la forma que este adopta, así como las prácticas sociales y los papeles que cada actor interpreta en la puesta en escena de un juicio “producen complejas interrelaciones entre texto y contexto” (Coulthard y Johnson, 2007: 68). Es frecuente, por ejemplo, que el intérprete proporcione una versión

“hiperformal del testimonio de los hablantes de las minorías lingüísticas”, alterando la percepción que los integrantes del jurado se forman sobre el nivel cultural y la capacidad intelectual del declarante (Ortega, 2011: 13).

Es, pues, de esperar que en los próximos años y como fruto de los esfuerzos que los profesionales de esta labor están realizando se elabore y aplique una normativa específica que la regule y, a la vez, valore adecuadamente el trabajo de quienes están preparados para realizarla¹¹.

Hemos querido incluir en este apartado, a modo de propuesta, la que consideramos un área de estudio novedosa: **el lenguaje policial**. Estrechamente vinculada al proceso judicial, la labor de las fuerzas y cuerpos de seguridad se apoya en el lenguaje para lograr sus objetivos de preservar derechos y libertades, por un lado, y llevar a cabo indagaciones tendentes a esclarecer delitos y faltas, por otro. Sin embargo, el habla policial, que en inglés ha venido a denominarse *policespeak* (Coulthard y Johnson, 2007: 76), no ha recibido la atención que merece, pues incluso los lingüistas forenses han centrado sus estudios en otras etapas del proceso judicial, como hemos visto.

Es cierto que hay algunos trabajos relevantes acerca de las entrevistas a víctimas o interrogatorios a sospechosos (Hall, 2008), pero como campo de estudio el habla policial ofrece un terreno mucho más amplio y todavía sin explorar. Recepción de llamadas telefónicas de aviso o socorro, entrevistas “de calle” con testigos, víctimas o sospechosos, transmisión por radio de información, órdenes e instrucciones, elaboración de informes verbales o escritos, comparecencias policiales o judiciales, comunicados, avisos a la población... todos ellos constituyen escenarios en los que la estrategia lingüística elegida influye decisivamente en los resultados obtenidos. El

¹¹ Es necesario destacar la labor que en este sentido realizan las asociaciones profesionales como la Asociación Profesional de Traductores e Intérpretes Judiciales y Jurados (APTIJ, www.aptij.es) o l'Associació de Traductors i Intèrprets Jurats de Catalunya (ATIJC, www.atijc.com)

discurso policial demanda de quien lo utiliza una competencia lingüística notable y en registros muy diferentes, con la dificultad añadida de que, con frecuencia, es preciso cambiar de uno a otro en muy breve espacio de tiempo. Testigos de un incidente, denunciantes, sospechosos de la comisión de un delito o compañeros de profesión, el trato con cada uno de ellos, a menudo reunidos en un mismo contexto espacial y temporal, requiere adaptar el discurso a sus *particulares* necesidades y características.

Por otro lado, la especial consideración de los policías, como agentes de autoridad y representantes del poder ejecutivo, confiere a su discurso una notable relevancia, que no siempre es tomada en cuenta, debido a la presión de situaciones que demandan una resolución inmediata. Pero incluso en esos momentos el ciudadano interlocutor u observador percibe con claridad la imagen que el agente transmite, no solo con sus actos, sino también con sus palabras. El lenguaje empleado va a definir en gran medida esa imagen, que a su vez influirá en aquella otra del Estado, a quien representan.

El uso que la policía hace del lenguaje se ajusta a la descripción del lenguaje-como-instrumento que hace Gail Stygall (1994). Los policías no solo manejan una jerga propia, sino que emplean el lenguaje de forma particular y adaptada a sus propias necesidades, constituyendo una auténtica comunidad de hablantes, tal y como John Gumperz ([1968] 2009) las describe. Nos parece, además, que los estudios realizados por Penelope Brown y Stephen Levinson (1987) en el campo de la interacción social, junto con su “Teoría de la Cortesía” y de los actos de amenaza o mantenimiento de las “caras positiva y negativa” de uno u otro interlocutor contribuirían a entender ciertos aspectos del habla policial. El policía debe salvaguardar en todo momento su imagen como agente de la autoridad, al tiempo que evita las amenazas a la imagen pública de las personas con las que trata, para lo cual debe recurrir a estrategias específicas en las que las relaciones de poder deberán estar siempre presentes.

Sin duda un estudio pragmático enriquecería el conocimiento lingüístico sobre una comunidad de habla poco sometida a estudio hasta el momento, a la vez que contribuiría a poner de manifiesto la importancia que el dominio del lenguaje tiene dentro del desarrollo de la labor policial.

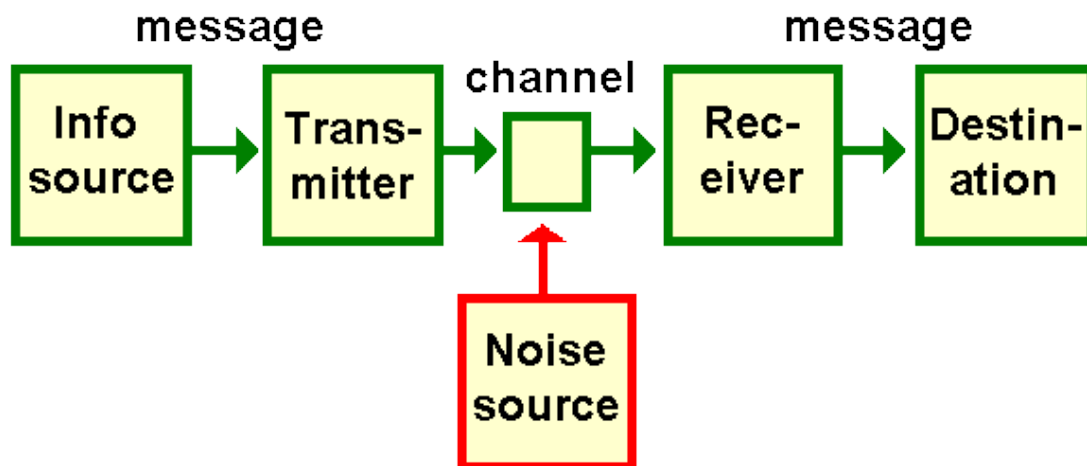
2.1.2.2. La atribución lingüística

Como hemos dicho más arriba, una de las dos grandes áreas de estudio de la lingüística forense se centra en la atribución de autoría de textos orales y escritos y, como veremos, en la consecución de este objetivo ha recorrido un trecho considerable en los últimos años. Sirviéndose de las herramientas que la técnica pone a su disposición, los lingüistas forenses se han concentrado en la tarea de diseccionar los textos objeto de estudio. Se trata de un trabajo arduo y laborioso, por mucho que las mencionadas herramientas lo simplifiquen en cierta medida, tanto que es posible que se hayan soslayado algunas cuestiones que suelen darse por sentadas.

Demostrar que una persona concreta ha dicho algo concreto, resulta, como reza el refrán popular, más fácil de decir que de hacer. Y esto es así porque el mero intento presupone que la palabra – en estado puro, despojada de los accidentes materiales que aportan los canales empleados – porta una seña, un vínculo, que la relaciona inequívocamente con el individuo que la formuló. Se trata de una premisa aventurada, casi fantástica, que merece, en nuestra opinión, una reflexión mayor de la que en los ensayos y monografías especializadas suele dedicársele.

En 1949 Claude Shannon y Warren Weaver, dos ingenieros de la compañía Bell Telephone, diseñaron un modelo del proceso de comunicación (*The Transmission Model of Communication*) del que han partido numerosos estudios sobre el tema. Este modelo, basado de hecho en la propuesta

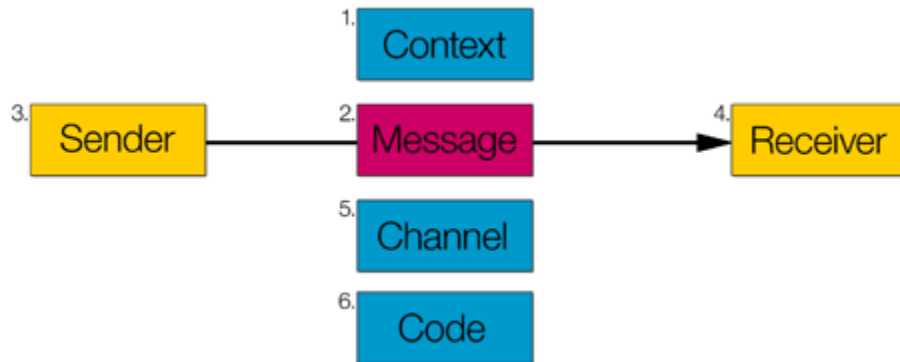
aristotélica de emisor, mensaje y receptor, consta de cinco elementos: una fuente de información de la cual parte el mensaje; un transmisor, que lo codifica; un canal, al cual se adaptan las señales transmitidas; un receptor, que descodifica el mensaje y un destinatario, a quien el mensaje va dirigido.



A estos elementos se le suele añadir un sexto: el *ruido*, todo aquello que distorsiona la señal, el mensaje, provocando una transmisión deficiente y dificultando el proceso.

Este modelo ha tenido una gran influencia y ha sido muy referenciado en estudios lingüísticos posteriores, en especial los que se han centrado en el proceso comunicativo, pero está, en la actualidad, considerado excesivamente mecanicista (Chandler, 1994) y desde su formulación ha sido revisado en numerosas ocasiones. Para James Carey (1989) se basa en una metáfora establecida durante el siglo diecinueve, la de que la información era “transportada” como podía serlo cualquier otra mercancía material. De acuerdo a la misma, las personas “colocan” ideas en el “interior” de las palabras, las cuales son llevadas por el canal elegido hasta el destinatario, quien “extrae” finalmente esas ideas y “absorbe” la información que contienen (Lakoff y Johnson, 1980: 10).

Roman Jakobson (1960: 353), a partir del modelo de Shannon y Weaver, estableció seis elementos que intervienen en el proceso comunicativo:



El esquema de Jakobson se ha convertido en una referencia en los estudios sobre el proceso comunicativo, pero ni en este ni el de Shannon y Weaver encontramos nada que sugiera que el mensaje porte en sí mismo un vínculo inequívoco con el emisor.

Sin embargo, esto no ha detenido a los lingüistas forenses, que han encontrado rastros que delatan la identidad del emisor en el *canal* utilizado y aquí nos estamos refiriendo a la voz del hablante y a los trazos dejados por el escritor manual. Tanto la *fonética forense* como la *grafología* han sido capaces de unir emisor y mensaje, si bien de una forma que podríamos decir indirecta, pues no ligan al hablante o al escritor con el contenido del mensaje, sino que recurren al elemento que lo soporta, los sonidos en un caso y los signos escritos en otro.

La *fonética forense* es, tal vez, el área de la lingüística forense de más rápido crecimiento y desarrollo, apoyándose para sus investigaciones en aparatos que recogen grabaciones fidedignas de la voz humana y en programas que la analizan y ofrecen estudios comparativos que señalan con

un muy alto grado de fiabilidad la identidad del hablante. Los análisis pueden ser fonéticos o fonológicos y descomponer los registros sonoros en gráficas asociadas a las características físicas que intervienen en la producción oral¹².

En palabras de Carlos Delgado:

“El habla es una referencia biométrica definida por la función y estructura anatómica de los órganos fono-articulatorios. También debe ser interpretada como un acto comunicativo y, por tanto, como algo dependiente de cada conducta individual” (Delgado, 2012: 101).

La *acústica forense* se centra en las características de los sonidos del habla, desde el punto de vista del emisor. Se trata de una ciencia altamente tecnificada que requiere de equipo especializado de registro del habla y un software apropiado para su procesamiento (McMenamin, 2002). Los avances en este campo en los últimos años han sido considerables y los numerosos estudios y abundante bibliografía disponible dan testimonio del grado de precisión alcanzado en sus análisis. En nuestro país destacan los trabajos realizados por Delgado (2001, 2005, 2012; Delgado et al., 2009), que recogen el estado actual de la acústica y fonética forense tal y como es puesta en práctica por el Laboratorio de Acústica Forense de la Policía española.

No sería aventurado asegurar que la Acústica Forense seguirá incrementando su relevancia dentro de las indagaciones policiales y judiciales en los próximos años. La creciente proliferación de elementos de comunicación a través de telefonía móvil e internet convierten el análisis de las conversaciones orales en una herramienta imprescindible para la

¹² El análisis de voz forma parte, igualmente, de otra disciplina, la *biometría*, que apoyándose en los más recientes avances tecnológicos busca la identificación de los individuos a partir de elementos corporales, como el reconocimiento automático a través del iris de los ojos, de patrones faciales o de la lectura óptica de las huellas dactilares. Su aplicación más inmediata va dirigida al desarrollo de sistemas de seguridad que protejan el acceso a instalaciones o equipos informáticos. Uno de los más recientes y peculiares sistemas de identificación, se basa en el análisis de patrones de movimientos al caminar, que investigadores de la Universidad Jaume I están desarrollando actualmente y que los medios de comunicación comparan, una vez más, con “la huella dactilar” (INNOVA, 02/08/13).

resolución de muchos casos. En España la colaboración entre la Universidad Politécnica de Madrid y la multinacional Indra ha propiciado la creación del Sistema de Identificación Biométrica Multimodal Aplicado a las Tecnologías de la Información (SIBMATI) (Garayzábal et al., 2012). Además, el sistema automático de reconocimiento de locutor *Batvox*, desarrollado en nuestro país por los profesores Javier Ortega y Joaquín González-Rodríguez, viene siendo empleado regularmente tanto por las fuerzas y cuerpos de seguridad como por empresas privadas¹³. El trabajo “La cualidad de voz en fonética judicial”, de Juana Gil y Eugenia San Segundo (2014), repasa de forma exhaustiva el estado de la cuestión en nuestro país, revisando la bibliografía más relevante y los protocolos más empleados en el análisis de grabaciones de voz: el *Voice Profile Analysis* y el denominado *GRBAS* (siglas en inglés de Grado de disfonía; Ronquera o aspereza; *Breathiness* o escape de aire; Astenia o hipofunción vocal; y *Strain* o tensión). Las autoras hacen un recorrido por la evolución de los sistemas de análisis de la cualidad de voz en las últimas décadas y describen algunos de los problemas con los que se encuentran los especialistas, como son la calidad de las muestras y la variación intra-locutor, problemas que, en general, resultan familiares a quienes abordan la atribución de autoría también a partir de textos escritos.

Por su parte la **grafología**, se apoya en los rasgos de la escritura manual para identificar al escritor. Curiosamente, la lingüística, que ha prestado una gran atención al soporte de la transmisión oral, no ha hecho lo mismo con el soporte de la transmisión escrita, la caligrafía. El estudio de la misma con el fin de identificar a su autor, la grafología, ha merecido en

¹³ La Biometría ha venido trabajando desde hace décadas en otra técnica sorprendente de atribución de autoría de textos escritos, si bien se trata en este caso de un elemento extralingüístico: los patrones de pulsado de teclas en el teclado del ordenador (*keyboard stroke dynamics* o *keystroke dynamics*). Al parecer, el ritmo con el que pulsamos las teclas es lo suficientemente característico como para identificarnos a cada uno de nosotros (Monrose y Rubin, 1999). La primera idea que nos viene a la mente es que dicho ritmo ha de variar en función de nuestro estado de ánimo, de salud, del nivel de atención y de muchos otros factores que deberían “engañar” al sistema y producir tanto falsos positivos como falsos negativos. Sin embargo, en la práctica, los numerosos sistemas comerciales diseñados para esta tarea han conseguido sortear con éxito dichos obstáculos y en la actualidad es uno de los sistemas de autenticación más utilizado en los países anglosajones (Lopatka, y Peetz, 2011).

ocasiones la misma desconfianza que se reserva para pseudociencias y para *actividades* tales como la astrología o la lectura de la palma de la mano. Tal vez el motivo se encuentre en su pretensión añadida de penetrar en la psique del escritor y emitir juicios acerca de la misma a partir de los trazos y rasgos de su escritura manual.

El origen de esta situación, que afecta sobre todo a nuestro país, radica en un malentendido, *en un error de traducción*, por así decirlo. Mientras que en España utilizamos el término *grafología*, de una forma general para referirnos a todo lo antedicho, en los países de habla inglesa se han cuidado mucho de separar el estudio lingüístico de la escritura o *grafémica*¹⁴, el estudio científico de la escritura manual o *grafonomía*¹⁵, y el examen forense de documentos o *documentoscopia*, de la *grafología* referida esta al estudio de la personalidad a través de la escritura manual. Tanto la grafonomía como la documentoscopia emiten informes que tienen plena validez jurídica y son solicitados y valorados con frecuencia por los tribunales a la hora de evaluar la autenticidad de un documento manuscrito¹⁶.

Fonética forense y grafonomía trabajan a partir de un mismo principio, el de igualdad o equivalencia, que Huber y Headrick, en su *Handwriting Identification: Facts and Fundamentals* (1999) definen como:

¹⁴ En un principio recibió el nombre de *gramatología*, por parte del lingüista que dio los primeros pasos en esta disciplina, Ignace Gelb (1952). El filósofo Jacques Derrida ([1967] 1997) tomaría, más tarde, prestado el término al elaborar sus teorías *deconstruccionistas*. Al parecer, en el momento de atribuir el nombre de *graphemics* a esta disciplina la primera intención era denominarla *graphology*, por lógica derivación de *phonology*, sin embargo esta denominación venía ya siendo empleada por la pseudo-ciencia que estudia el carácter del escritor a través de la escritura y hubieron de conformarse con el de *graphemics*.

¹⁵ La *grafonomía*, se ocupa igualmente del estudio de la escritura manual, con una visión interdisciplinar y reúne en el ejercicio de su tarea conocimientos de áreas tan diversas como la paleografía, la neurociencia o el reconocimiento automático de la escritura manual. Los especialistas en grafonomía se agrupan en la International Graphonomics Society (IGS) y en su página web proporcionan una breve definición del término: “Graphonomics is the multi-disciplinary field of fundamental and applied experimental research of handwriting and related skills.”

¹⁶ En el momento de escribir estas líneas, el caso más reciente y de mayor difusión mediática es el llamado “Caso Bárcenas”, donde el juez instructor solicitó la peritación de diversos documentos contables con el fin de atestiguar su autoría. Aquí, una vez más los medios de comunicación se refirieron a “un estudio grafológico” (ABC, 06-07-2013; El Mundo, 16-07-2013; El País, 30-04-2013).

"When any two items possess a combination of independent discriminating elements (characteristics) that are similar and/or correspond in their relationships to one another, of such number and significance as to preclude the possibility of their occurrence by pure coincidence, and there are no inexplicable disparities, it may be concluded that they are the same in nature or are related to a common source." (Huber and Headrick, 1999: 84).

Sin embargo, tanto en el caso de la fonética forense, como en el de la grafonomía, y aun reconociendo sus méritos, es preciso admitir que trabajan sobre el *canal* por el que se transmite el mensaje, no sobre el *mensaje* en sí mismo. Es decir, que tanto para quien analiza la voz como para quien analiza la escritura, el contenido, la forma de elaborar el mensaje, es indiferente. Por el contrario, la estilística forense y la estilometría toman como materia de estudio y análisis el mensaje y la forma en que está redactado. La hipótesis de la que parten establece que cada individuo se expresa – ya sea oralmente o por escrito – de forma única y personal, haciendo un uso del lenguaje tan idiosincrásico que es posible vincular un enunciado con la persona que lo formuló.

Esta hipótesis de trabajo permite a los lingüistas analizar textos escritos y determinar si el autor ha incurrido en plagio o atribuir un texto dubitado a un escritor concreto. Ambas tareas parten de presupuestos similares y se apoyan en los fenómenos de la variación lingüística y del idiolecto.

2.1.3. Variación lingüística e idiolecto. El soporte teórico de una hipótesis de trabajo

Lo cierto es que ya F. de Saussure, en su *Curso de Lingüística General* ([1916] 2008), sugiere una producción tan diversa – al menos en teoría – como el número de hablantes, en la distinción que efectúa entre *langue*, o lengua, el sistema estructurado de reglas y convenciones gramaticales, y *parole*, el habla, la puesta en práctica, la materialización por parte del individuo de ese sistema que ha interiorizado como miembro de la comunidad lingüística a la que pertenece.

“El ‘habla’ es la suma de todo lo que la gente dice, y comprende las combinaciones individuales, dependientes de la voluntad de los hablantes y los actos de fonación igualmente voluntarios, necesarios para ejecutar tales combinaciones” (Alcaraz y Martínez, 1997: 332).

El concepto clave parece ser el de la “voluntad” del hablante, que también es mencionado como relevante por Lázaro Carreter en su interpretación de *langue*:

“El habla es un acto individual de voluntad y de inteligencia, en el cual conviene distinguir:

- 1) Las combinaciones por las que el sujeto hablante utiliza el código de la lengua con miras a expresar su pensamiento personal.
- 2) El mecanismo psicológico que le permite exteriorizar esas combinaciones. El habla es la suma de todo lo que las gentes dicen y comprende:
 - a) Combinaciones individuales, dependientes de la voluntad de los hablantes.
 - b) Actos de fonación igualmente voluntarios, necesarios para ejecutar tales combinaciones” (Carreter, 2008: 217).

Y por Alcaraz Varó, para quien las combinaciones individuales son “dependientes de la voluntad de los hablantes”, así como “los actos de fonación igualmente voluntarios, necesarios para ejecutar tales combinaciones” (Alcaraz y Martínez, 1997: 322). Es este ejercicio de voluntad el que permite al individuo, en la práctica, elegir, seleccionar y aplicar toda una serie de preferencias que convierten sus enunciados en ejemplos diferenciados de los de otros individuos con los que comparte una lengua determinada.

El estructuralismo sausseriano influyó el pensamiento de la Escuela de Praga, uno de cuyos fundadores, Roman Jakobson, desarrollaría la idea de la articulación variable por parte del hablante en el concepto de “literariness”, o la capacidad de manipular el lenguaje para construir una forma de expresión diferenciada, que, a su vez, da como resultado una obra de literatura (Baldick, 2001). Jakobson establece la diferencia entre el lenguaje cotidiano y el lenguaje literario, atribuyendo a este último una serie de características y propiedades formales que lo distinguen del primero. No es difícil concluir que para hacerlo, el hablante – o escritor – debe manipular el lenguaje de acuerdo a sus preferencias y capacidad.

El lingüista inglés M. A. K. Halliday desarrollaría esta visión de la capacidad que el usuario de una lengua posee para manipularla, si bien desde posiciones alejadas del estructuralismo. Halliday vincula la forma que adopta el lenguaje al contexto en el que se desenvuelve y utiliza el término *registro* para referirse a las distintas opciones que se le ofrecen al hablante, opciones influenciadas por el tema tratado, el tono de la conversación y por las personas que intervienen en la misma (Halliday, 1985). Concibe la gramática como sistémica y funcional, dependiente del entorno “eco-social” y otorga un elevado grado de control al hablante sobre el lenguaje.

Esta “gramática sistémica-funcional” implica que cada usuario de una lengua tiene a su disposición un conjunto de elecciones posibles que se estructuran jerárquicamente (Mathews, 1997). Sin embargo, toda gramática, asegura Halliday, describe los sistemas de una lengua, las relaciones entre ellos y la forma en que se ejecutan, “hasta un nivel de detalle tal que las elecciones que restan se limitan a conjuntos abiertos de unidades léxicas” (Mathews, 1997: 370).

Esta disposición podría parecer ciertamente restrictiva y también André Martinet parece sugerir en un primer momento una idea similar cuando afirma que “la originalidad del pensamiento no se podrá manifestar más que con una disposición inesperada de las unidades” (Martinet, [1960] 1984: 23). A pesar de ello, con su propuesta de la doble articulación del lenguaje – a la vez economía y diversidad de producción – Martinet deja claro que las posibilidades que ofrece el lenguaje son en realidad tan amplias que, en la práctica, dan origen a infinidad de estilos y formas de expresión claramente diferenciados entre sí¹⁷. Esta paradójica combinación de elementos restrictivos y libertad creativa es resaltada por Eugenio Coseriu:

“El sistema es sistema de posibilidades, de coordenadas que indican caminos abiertos y caminos cerrados: puede considerarse como conjunto de “imposiciones”, pero también y quizá mejor, como conjunto de libertades, puesto que admite infinitas realizaciones y solo exige que no se afecten las condiciones funcionales del elemento lingüístico: más bien que “imperativa”, su índole es consultiva” (Coseriu, 1973: 98).

¹⁷ Más tarde, sin embargo, Martinet admite, como no puede ser de otro modo, que el hablante está permanentemente eligiendo, seleccionando entre las diversas opciones que el lenguaje lo ofrece para transmitir sus experiencias y reconoce que “la realidad es que, en las relaciones de una misma comunidad, la absoluta identidad de los sistemas parece ser más bien una excepción” (Martinet, [1960] 1984: 180).

“La labor espiritual del individuo hablante consiste justamente, en la aplicación original del sistema, dentro y fuera de lo permitido por la norma...” (Ibid, 99).

Esta libertad creativa, esta “aplicación original del sistema, dentro y fuera de lo permitido por la norma”, como afirma Coseriu, da en la práctica origen a variedades individuales o idiolectos, tan reales como difíciles de sistematizar y en su existencia se basa la estilística forense.

Diversidad y variación son inherentes al ejercicio lingüístico y constituyen la premisa de trabajo sustentada por la Teoría de la Variación Lingüística que William Labov formuló hacia mediados del siglo pasado y que establece que aun compartiendo el mismo código lingüístico, los hablantes de una lengua determinada producirán, en la práctica, estructuras superficiales distintas para una misma estructura profunda (Labov, [1966] 2006). Una lengua es compartida por una comunidad amplia, dentro de la cual existen subgrupos sociales que utilizan su propio dialecto y dentro de estos cada individuo hace uso de los recursos idiomáticos de manera personal. Esta división se puede multiplicar repetidas veces, dado que los subgrupos lingüísticos pueden fragmentarse a su vez en otros de menor tamaño. Dentro de la comunidad de hablantes de español, es posible distinguir en primer lugar las distintas nacionalidades de origen y a su vez dentro de estas, las diferentes regiones geográficas. Además, un hablante de nacionalidad española, que posea ciertos rasgos dialectales debidos a su origen levantino, por ejemplo, puede acumular otros en función de su adscripción a determinados grupos sociales o culturales y emplear, de acuerdo a las circunstancias, un registro lingüístico u otro en un ejercicio diglótico que sumará rasgos diversos a su producción final.

El fenómeno de la variación ha supuesto a menudo un problema para quienes buscan formular teorías que describan y expliquen de forma general

la producción lingüística, lo que les ha llevado a minimizar su importancia o incluso a ignorarlo por completo (Honeybone, 2011). Resulta comprensible que a la hora de formular reglas o normas que describan un proceso en su aspecto más amplio, las excepciones tiendan a ser obviadas, calificándolas de irrelevantes. Así, Chomsky (1965) lidia con esta dificultad alegando que las variaciones son debidas a factores extralingüísticos y que un cierto grado de idealización es necesario si se desea encontrar los patrones que regulan el uso de una lengua¹⁸. Sin embargo, Labov ([1972] 1991), al establecer los fundamentos de la sociolingüística, abrió la puerta a posteriores estudios de las causas y los efectos de la variación lingüística. Esta variación acostumbra a clasificarse en dos grandes grupos: la que tiene lugar entre los hablantes de una misma lengua, o *inter-hablantes*, y aquella otra que afecta a la producción de un solo hablante, o *intra-hablante*. Ambos fenómenos coexisten en una misma realidad lingüística y han sido objeto en los últimos años de abundante investigación (Adger, 2006; Bender, 2007; Lodge, 2009; Vaux, 2008). Chomsky (1986) propone una distinción entre el conocimiento acumulado por el hablante que él denomina *I-language* y que da lugar a aquellas manifestaciones verbales que proceden del *i-nterior* del *i-ndividuo* y *E-language*, referido a todas aquellas producciones que se ven afectadas por factores *e-xternos* al lenguaje, como la comunidad a la que el hablante pertenece, por ejemplo. Por esta razón la teoría lingüística de Chomsky (1965) se ocupa de un hablante oyente ideal, ubicado en una comunidad de habla por completo homogénea.

Quienes se centran, pues, en encontrar una explicación a todas aquellas variaciones que se producen entre los hablantes de una misma lengua optan por una aproximación *panlectal* (Bailey, 1973), mientras que en el extremo opuesto estarían quienes eligen estudiar las variaciones generadas por un solo

¹⁸ Martinet también reconoce los problemas que la variación supone para el lingüista descriptivo cuando asegura que “nos hemos apresurado a olvidar estas diferencias para no complicar la exposición; el análisis de una lengua a la que se da como uniforme es algo tan delicado, que conviene simplificar al máximo los datos del problema” (Martinet, [1960] 1984: 185).

hablante o *idiolecto* (Honeybone, 2011). Ambas opciones resultan de interés para el lingüista forense, más preocupado por la información que las manifestaciones verbales anónimas objeto de estudio puedan proporcionar acerca del origen y la identidad de su autor, que en encontrar una explicación para dichas variedades. La procedencia geográfica del hablante-escritor, su formación cultural y académica, su pertenencia a determinados grupos sociales y su género serán datos de gran importancia, pero sobre todo lo serán aquellos rasgos lingüísticos que permitan individualizarle de entre un grupo más o menos numeroso de autores-candidatos.

El lingüista forense está convencido de que es posible describir y medir la variación lingüística y de que analizando el origen de los cambios que el autor introduce en su producción oral o escrita es posible establecer una serie de rasgos característicos que serán de gran ayuda y contribuirán significativamente en la indagación judicial de la que hablábamos más arriba. Las variaciones fonéticas, léxicas o sintácticas pueden tener su origen tanto en factores internos – lingüísticos – como externos – extralingüísticos – pero en cualquier caso son susceptibles de ser sistematizadas y clasificadas como valiosas herramientas de análisis estilístico.

Esta aproximación a la variación lingüística surge en la década de 1960 (Coulthard, 2007) y desde entonces este fenómeno ha recibido una atención cada vez mayor dentro de los distintos estudios teóricos, que buscan integrarlo en los sistemas descriptivos generales.

“A natural language cannot be successfully observed outside a theoretical paradigm, but the paradigm cannot be constructed without the observation of language as it is used.” (Coulthard, 2007: 65).

Encontrar un patrón que regule e incluso pueda predecir estos cambios, es crucial en el establecimiento de una disciplina que busca obtener

conclusiones prácticas a partir del estudio de los mismos, como es el caso de la lingüística forense. Si la población de una determinada región de España introduce en su habla cotidiana ciertas variaciones fonéticas, que no se dan en ninguna otra región, y dichas variaciones son constantes y regulares, servirán para atribuir un origen geográfico al hablante objeto de estudio.

El orden y la regularidad, por lo tanto, van a ser de una gran importancia en el análisis forense del lenguaje. Si un individuo determinado introdujese variaciones aleatorias en su particular idiolecto, sería poco menos que imposible atribuirle una muestra textual, incluso aunqueuviésemos la certeza de antemano de que ha sido producida por él, ya que careceríamos de un “perfil lingüístico” característico de dicho individuo. Por fortuna los estudios realizados en los últimos años muestran que no es así; el lenguaje, asegura McMenamin (2002), no tiene por qué carecer de variación (polimorfismo) para ser estructurado, ni necesita ser homogéneo para ser regular.

“Nativelike command of heterogeneous structures is not a matter of multidialectalism or ‘mere’ performance but is a part of unilingual linguistic competence.” (Weinrich, Labov y Herzog, ([1968] 2013).

“Every man’s language has, first, its individualities; second, the common properties of the class to which he belongs; and third, words and phrases of universal use.” (Coleridge, [1817] 2012).

El idiolecto estará, entonces, formado por la suma de todos estos rasgos lingüísticos más aquellos propios del individuo: sus características físicas que imprimirán a su voz un timbre característico y sus procesos cognitivos que condicionarán la disposición sintáctica de sus enunciados, así como su formación cultural determinará la riqueza léxica. McMenamin (2002: 71), asegura que el idiolecto “es el resultado de una combinación única,

efectuado de forma inconsciente¹⁹, de conocimientos lingüísticos, asociaciones cognitivas e influencias extralingüísticas”. Es el idiolecto el que lleva al autor de una nota de petición de rescate, por ejemplo, a incluir elementos que permitirán a los lingüistas forenses obtener información relevante para su identificación en un primer momento y para su inculpación ante los tribunales, más tarde.

2.1.4. Paradojas y aporías

Parece, pues, suficientemente demostrada la libertad que el lenguaje otorga a los usuarios, a pesar de la aparente restricción que en principio suponen las estructuras gramaticales. Los hablantes y escritores disponen, sino de una libertad absoluta, sí de una *libertad recursiva*, la cual, mediante la acumulación repetida de una serie de opciones restringidas, se materializa en la práctica en la posibilidad de realizar un número infinito de combinaciones.

Sin embargo, si el hablante llevase hasta las últimas consecuencias esta libertad de expresión, ello sería igualmente contraproducente para el lingüista forense, ya que si por un lado precisa que el autor cuyos textos somete a estudio se exprese de forma distinta al resto de los usuarios de su misma lengua; precisa al mismo tiempo que este repita de forma regular una serie de elementos reconocibles. Si el cambio es deseable en lo relativo a su relación con la comunidad de hablantes, también lo es que mantenga estables los rasgos que servirán para identificarle en relación consigo mismo. El hablante ideal, para el lingüista forense, es a la vez distinto y constante en su expresión verbal. De no ser así, sería indistinguible de sus coetáneos lingüísticos en el primer caso e irreconocible con respecto a sí mismo en el segundo.

¹⁹ Este detalle resultará de suma importancia en el análisis forense y le prestaremos atención posteriormente. Adelantaremos, sin embargo, que identidad y producción verbal inconsciente se hayan estrechamente vinculadas. Cuanto más reflexivo sea un texto, cuanto mayor haya sido el cuidado en la elección de las palabras y en su disposición sintáctica, más difícil será para el lingüista encontrar en él rasgos idiosincrásicos comunes a toda su producción lingüística.

Al plantearnos la tarea del examen estilístico forense se dan algunas paradojas generadas por la necesidad de atribuir tanto al lenguaje como al hablante/escritor cualidades opuestas. Por un lado, la estructura gramatical dentro de la que se desenvuelve el individuo restringe las opciones disponibles, pero al mismo tiempo proporciona la libertad suficiente como para producir variaciones inacabables. Por otro, estas variaciones son, como nos aseguran los lingüistas mencionados más arriba, fruto de un control volitivo que el hablante/escritor ejerce sobre el sistema y sobre su producción lingüística. Sin embargo, y al mismo tiempo, en la producción de ese mismo hablante/escritor, interviene – debe intervenir – una parte ajena a su voluntad, que le lleva a repetir, una y otra vez, en distintos textos, los mismos rasgos y características que el lingüista forense utilizará para atribuirle un texto dubitado. Resulta difícil concebir un proceso de escritura que conjugue una selección léxica y sintáctica consciente y a la vez inconsciente, pero de hecho así parece ser, a tenor de los innumerables estudios realizados y que sustentan la atribución de autoría en su presente estadio.

Se trata, sin duda, de un problema, de un fenómeno excepcional, que debe ser estudiado en profundidad a partir de la abundante documentación que los trabajos de atribución de autoría están generando y ante el que la lingüística forense ha optado por una aproximación pragmática; aceptando la realidad tal y como es, sin esperar a una teoría general que la explique.

Ahora bien, aun dejando de lado estas posibles aporías, sigue existiendo otro problema, tan real como obviado, el de hasta qué punto un hablante/escritor emplea una versión única e irrepetible de la lengua. La cuestión fundamental no sería tanto el hecho de si efectivamente cada hablante produce o no su propia versión de la lengua en que se expresa – algo que parece suficientemente probado – sino hasta qué punto esa producción es única. Nos parece que, en ocasiones, partiendo de una premisa cierta – la diversidad de la producción lingüística individual – se ha extraído una

conclusión demasiado atrevida: la de que cada idiolecto es irrepetible. Si lo hacemos así caemos de nuevo, en nuestra opinión, en un error de método, pues que diez, cien o mil escritores tengan un estilo propio y diferenciado, no es base suficiente para aseverar que no existen dos – en algún lugar – que tengan un estilo, sino idéntico, sí al menos tan similar que sea imposible discriminar entre ellos²⁰.

Y surgen todavía otras cuestiones, sino problemáticas, sí, al menos, inquietantes. Por ejemplo: Aceptando la existencia *de facto* de esos rasgos lingüísticos que, al parecer, de forma inconsciente imprimimos en nuestros textos, ¿no deberíamos interpretar esa realidad como una evidencia de un cierto determinismo lingüístico²¹? ¿No significa acaso su presencia que estamos condicionados a hablar/escribir como lo hacemos? Si a pesar de toda nuestra libertad expresiva dejamos siempre un “rastros” lingüístico en nuestros enunciados, nos encontramos ante un condicionamiento sobre cuyo origen solo podemos, por el momento, especular. Hasta dónde llega este condicionamiento y a qué elementos léxicos y sintácticos afecta es precisamente lo que la estilística forense intenta determinar para llevar a cabo su tarea.

Por otro lado, los lingüistas forenses acostumbran a asegurar que la “huella lingüística” del hablante/escritor es irrepetible, y en ello se basan para atribuir un texto a una persona concreta una vez cotejadas la “huella” y el “perfil” del autor. Sin embargo, cabe preguntarse si no sería posible reproducir esa “huella”, una vez que hemos obtenido el “perfil” que identifica al autor. Los rasgos que caracterizan a un autor determinado y en los que se basa la estilística forense son elementos concretos y mensurables y, por tanto,

²⁰ Lo cierto es que a partir de cierto punto la generalización deviene en regla y los casos que se apartan de ella se convierten en excepciones: a estas alturas importaría bien poco si en algún momento y lugar apareciesen dos personas que tienen dos huellas dactilares idénticas, puesto que su carácter de excepcionalidad no invalidaría en absoluto el método de identificación por medio de la dactiloscopia.

²¹ Se trataría de un determinismo lingüístico por completo diferente del que considera que nuestra visión del mundo en que habitamos está condicionada por el lenguaje que utilizamos para describirlo y que viene a ser conocido como la Hipótesis Sapir-Whorf (Deutscher, 2010).

reproducibles. Nada nos impide, al menos en teoría, emplear los elementos que configuran la identidad estilística de un escritor determinado – Baroja, Faulkner, Cela, Proust – y escribir un texto que, en rigor debería serles atribuido a ellos. Por otro lado, al reproducir la “huella” de otro escritor, de tal manera que nuestra producción textual imite a la suya, ¿no estaríamos contradiciendo ese supuesto condicionamiento del que hablábamos más arriba? Si dejo de escribir como “yo mismo” y lo hago como “otro” estaré demostrando con los hechos que no existe el referido determinismo.

Y todavía otra cuestión más: Suponiendo que pudiésemos reproducir el estilo de los escritores mencionados, ¿tendrían acaso nuestros productos, las “réplicas”, el mismo valor que los textos producidos por las manos de los escritores “originales”? La cuestión puede sintetizarse en la pregunta: ¿si una imitación es indistinguible del original no debería ser tan valiosa como este? La tentación más inmediata es responder que no²², pero ¿en qué nos basamos?

Se trata, en realidad de preguntas engañosas, viciadas en su mismo origen, pues definir el estilo de escritura de Baroja, de Faulkner, de Cela, de Proust – o de cualquier otro escritor – a partir del conjunto de marcas empleadas para atribuirles la autoría de un texto peca de reduccionismo. Su estilo es “eso”, pero no “solo eso”. La construcción de la historia, de los personajes, del ambiente y el uso de las figuras de estilo – metáforas, símiles, contraposiciones, clímax, elipsis, etc. –es en realidad lo que proporciona la sensación inequívoca de estar leyendo a uno de estos escritores. Los motores de búsqueda actuales no alcanzan por el momento a aislar estos aspectos, cruciales en la formación de la identidad literaria de un autor, pero los avances de la lingüística computacional son tales y tan veloces que es arriesgado incluso predecir de qué serán capaces en un futuro próximo.

²² Nos deslizamos aquí por un terreno resbaladizo. Los recientes avances en genética han provocado un extenso, bien conocido y en ocasiones encendido debate entre posiciones contrapuestas sobre esta misma cuestión, solo que trasladada al plano más trascendente de la vida humana.

Cuando llegue ese momento – si finalmente llega – estas preguntas estarán plenamente vigentes y sin duda darán lugar a interesantes debates.

2.2. Estilometría: la estadística al servicio de la lingüística

La Estilometría se apoya en técnicas cuantitativas para analizar el estilo de escritura, habitualmente con la finalidad de determinar la autoría de un texto dubitado o detectar un posible plagio (Garayzábal, 2012; Kenny, 1986; McMenamin, 2012). En esta tarea la Estilometría hace uso de técnicas estadísticas que varían de acuerdo a las preferencias del investigador y de las características del texto cuestionado (Stefanova, 2009). Para Chaski (2005) la Estilometría ofrece una aproximación computacional al problema de la determinación de autoría enfocado en aquellos rasgos lingüísticos susceptibles de ser cuantificados, como pueden ser la longitud de las palabras, de las clausulas o de las frases, la frecuencia léxica o la distribución de palabras de distintas longitudes. Anthony Kenny, en su estudio del *Nuevo Testamento*, afirma que la Estilometría es “el estudio de los rasgos cuantificables del estilo de un texto oral o escrito” (Kenny, 1986: 1) y para Daniel Pavelec, “la estilometría ayuda a definir los rasgos discriminativos, como la riqueza de vocabulario y la repetición léxica a partir de la frecuencia de aparición” (Pavelec et al., 2009: 670).

2.2.1. Estilometría y Estilística Forense: Dos aproximaciones a los problemas de estilo e idiolecto

Proporcionar una definición accesible de esta rama de la lingüística forense, como puede verse, no es difícil. Sin embargo, una segunda lectura más atenta revela aspectos que pueden generar alguna incertidumbre. La

Estilometría pretende, en efecto, estudiar o más bien analizar el particular estilo de escritura que cada individuo posee, aunque ya veíamos en la Sección anterior que esta tarea puede resultar problemática. Si intentamos determinar con precisión en qué consiste un “estilo” de escritura o habla – más allá de los rasgos grafémicos o fonológicos – la tarea comienza a complicarse, sobre todo si pretendemos mostrar estas diferencias de una forma objetiva y cuantificable. Para Olsson (2008: 29), *estilo* es un término cuyo uso ha sido sometido “a todo tipo de excesos, tanto en estudios lingüísticos, como literarios”.

Hasta hace pocos años, la mayoría de los estudios comparativos se basaban en una aproximación cualitativa y se apoyaban en análisis estilísticos. Como disciplina vinculada tanto a la literatura como a la lingüística, la *estilística* se ocupa del estilo en el uso del lenguaje.

“Stylistics has been defined as a sub-discipline of linguistics that is concerned with the systematic analysis of style in language and how this and vary according to such factors as, for example, genre, context, historical period and author.” (Jeffries y McIntyre, 2010: 1).

Si nos atenemos al objeto de la Estilometría, esto es, determinar la autoría de un texto dubitado, y nos planteamos la tarea desde un punto de vista *estilístico*, precisaríamos obtener los rasgos que constituyen la esencia del “estilo literario” tanto del texto en cuestión como de los textos escritos por el o los autores candidatos, a fin de realizar un análisis contrastivo. La estilística busca, pues, la determinación de los rasgos formales, pero su objeto es habitualmente la interpretación del texto en cuestión (Wales, 1989; Simpson, 2009). La estilística, cuyo origen se remonta a la escuela de crítica literaria formalista que surgió a principios del siglo XX en Rusia, se apoya en modelos y técnicas de análisis lingüísticos con el fin de examinar sobre todo

el estilo literario, aunque también ha prestado atención a otro tipo de textos no-literarios.

Ciertos lingüistas, como en el caso de Gerald McMenamin, han desarrollado las técnicas de análisis estilístico y las han aplicado al contexto forense dando lugar a la “estilística forense”.

“Stylistics exploits the two principles of inherent variability in language: no two writers of a language write in exactly the same way, and no individual writer writes the same way all the time. Forensic stylistic analysis makes use of stylistic analysis (stylistics) to reach a conclusion and opinion related to the authorship of a questioned writing within the context of litigation. Stylistics is the scientific study of patterns of variation in written language. The object of study is the language of a single individual (idiolect), resulting in a description of his or her identifying linguistic characteristics.” (McMenamin, 2002: 176).

Basándose sobre todo en la Teoría de la Variación Lingüística, McMenamin busca reconocer los patrones de cambio que se producen *entre* los individuos, al tiempo que intenta identificar rasgos idiolectales *dentro* de un individuo aislado, lo que en cierta manera entra en contradicción con ese segundo principio de la variación lingüística mencionado, al principio de la cita, de que nadie escribe siempre de la misma forma.

Si la variación lingüística presenta un problema para los lingüistas en general, el *idiolecto* representa un problema muy concreto para los lingüistas forenses en particular, a pesar de que de su existencia ya nadie duda a estas alturas:

“Como es bien sabido, no hay dos personas que hablen o escriban igual, sino que cada uno utiliza un “estilo” de hablar o escribir

idiosincrásico. Este estilo es llamado idiolecto” (Cicres y Turell, 2012: 185).

“Even within a narrowly defined dialect community, individuals will have their own preferred detailed pronunciations of particular words. The combination of number of such preferred alternative pronunciations yields an overall pronunciation which is idiosyncratic, [and] that is, an individual’s idiolect.” (Nolan, 1994).

“The linguist approaches the problem of questioned authorship from the theoretical position that every speaker has their own idiolect, and the assumption that this idiolect will manifest itself through distinctive and idiosyncratic choices in text.” (Coulthard, 2004: 431).

Para Cicres y Turell el elemento sobre el que se apoya la investigación lingüística forense es la presuposición de que todas las personas utilizan el lenguaje de un modo distinto e idiosincrásico (2012). El problema del que hablamos más arriba radica en el hecho de que el habla o la escritura de cada individuo no permanece constante a lo largo del tiempo, a diferencia de otros rasgos biométricos como las huellas dactilares o el dibujo de las retinas. El propio William Labov, autor de la Teoría de la Variación Lingüística a la que recurren con frecuencia los lingüistas forenses, si bien reconoce la existencia del *idiolecto*, añade que nadie “parece haber sido capaz de encontrar un patrón homogéneo en el mismo” (Labov, [1972] 1991: 192).

A pesar de ello, a pesar de que nuestros patrones lingüísticos cambian con el tiempo, debido a muchas circunstancias – el nivel cultural, el medio social en el que nos desenvolvemos, el tema sobre el que hablamos – los estudios forenses del habla y la escritura precisan que algo permanezca constante en nuestra producción verbal y escrita. De lo contrario, sería imposible asignar un perfil lingüístico a cada individuo y todo el proceso de atribución de autoría se vendría abajo. Un investigador confía – porque la

práctica así se lo ha demostrado – en la existencia de marcadores de estilo (McMenamin, 2002) que le permitirán contrastar textos dubitados e indubitados. La mecánica que regula a la vez el cambio y la constancia en el *idiolecto* no ha sido por el momento suficientemente estudiado, a pesar de lo cual la lingüística forense continúa adelante con su investigación, en un excelente ejemplo de que práctica y teoría no tienen por qué avanzar a un ritmo sincronizado²³.

Así pues, es el *idiolecto* el elemento fundamental en torno al que gira la tarea de atribución de autoría y respecto a cuál puede ser el mejor medio para medir los rasgos que lo integran existen múltiples y divergentes opiniones:

“El índice de similitud idiolectal (ISI) es un índice que mide el grado de similitud entre el idiolecto que se extrae de dos o más muestras lingüísticas, orales o escritas. Este índice es un valor comprendido entre el 0 y el 1, de modo que el 1 indica que los idiolectos analizados son idénticos, y el 0 que son muy lejanos. A efectos forenses, la hipótesis que interesa evaluar es que un ISI con un valor bajo es el que permitiría discriminar mejor a los hablantes, mientras que un ISI con valor cercano a 1 permitiría identificarlo”. (Cicres y Turell, 2012: 193).

La estilística forense elige una aproximación cualitativa, centrando la búsqueda de marcadores de estilo en errores o incorrecciones léxicas y gramaticales, y en rasgos inusuales o no estándar, que generalmente suelen ir asociados a dialectos regionales o comunidades del habla concretas (Olsson, 2008). La información que estos rasgos proporcionan, incluso aunque en ocasiones no resulten suficientes para una identificación plena, suelen ser de gran utilidad en investigaciones penales para determinar la procedencia, estatus y formación del autor del texto, datos que contribuyen de forma

²³ La psicolingüística, sin embargo, ha venido realizando trabajos muy interesantes en este campo, vinculando el empleo de ciertos términos y vocablos a procesos cognitivos ajenos a la voluntad del individuo, lo que se ajusta a la perfección a las necesidades forenses de los lingüistas que intentan encontrar marcadores de estilo estables. (Garrett, 1982).

indirecta a su localización y posterior identificación. El perfil elaborado mediante un análisis estilístico forense posibilita a los investigadores descartar a determinados candidatos, reduciendo así el número de sospechosos en los que centrar las indagaciones.

Los resultados de un análisis cualitativo de un texto, basado en un estudio estilístico forense se apoyan, en gran medida, en la habilidad, experiencia y conocimientos del investigador por lo que en ocasiones carecen del grado de objetividad y precisión que proporciona un análisis cuantitativo. Carol Chaski, en su defensa de los métodos estilométricos, asegura que para ser efectiva, la aproximación estilístico-forense precisa de extensas bases de datos poblacionales de rasgos lingüísticos, que soporten la relevancia de los rasgos estilísticos señalados, algo que hoy por hoy no existe. Sin estas bases de datos, añade Chaski, “la intuición del investigador acerca de dicha relevancia puede conducir a conclusiones subjetivas y sesgadas” (Chaski, 2005).

La Estilometría, por otro lado, al trabajar sobre elementos cuantificables medidos y procesados por sistemas informáticos aporta la objetividad y precisión propias de un estudio cuantitativo, al menos en teoría. En la práctica ya veremos que no es exactamente así, pues la decisión crucial de qué rasgos lingüísticos van a ser empleados como elemento discriminador sigue siendo decisión del investigador, por lo que el elemento humano, con su carga inherente de subjetividad, continúa presente en proceso, aunque sea en menor medida.

El uso de herramientas estadísticas para resolver problemas en la autoría de textos escritos surge hacia mediados del siglo XIX. El matemático Augustus de Morgan sugirió en 1851 vincular la longitud media de las palabras con la identidad de un autor determinado y hacia 1887, el físico Thomas Mendenhall desarrolló esta misma medida y la frecuencia relativa de

aparición de determinadas palabras para elaborar una “curva característica”, propia de cada escritor²⁴ (Koppel, Schler y Argamon, 2009). Otro matemático, William Benjamin Smith, por su parte recurrió a la longitud media de las frases en su estudio de las Epístolas de San Pablo (Schaalje, Roper y Fields, 2012; Holmes, 1998).

Es el polaco Wicenty Lutoslawski, sin embargo, quien pasa por ser el fundador de la *estilometría*, o al menos quién acuñó el término y definió los principios de esta, por aquel entonces, nueva ciencia²⁵. En 1890 publicó el tratado teórico, *Principes de stylometrie* y más tarde llevó a cabo un estudio con el que estableció la cronología de los *Diálogos* de Platón, en donde señala los distintos tipos de palabras, frecuencia de aparición, disposición y otras marcas estilísticas, de las que llega a reunir quinientas (Lutoslawski, [1897] 2006).

En la década de 1930 el lingüista George Kingsley Zipf demostró que existe una relación constante e inversa entre el número de veces que un ítem léxico aparece en un texto y su posición relativa en la lista de frecuencia de aparición, lo que hoy es conocido como la Ley de Zipf (Holmes, 1998). Basándose en esta regla, el inglés George Udny Yule intentó establecer una medida objetiva de la frecuencia de aparición de una palabra, a la que se ha dado el nombre de “*característica K de Yule*”, que más tarde se revelaría ineficaz como método para discriminar autoría literaria. Yule exploró igualmente las posibilidades de la longitud media de las frases, para concluir que tampoco era un parámetro fiable (Koppel, Schler y Argamon, 2009).

²⁴ Mendenhall realizó también estudios sobre la posible autoría de Bacon de parte de las obras de Shakespeare en los que tomó como referencia la longitud de las palabras, que contaba por medios puramente visuales y manuales. (Coulthard y Johnson, 2007: 164). Para David I. Holmes, “His legacy was to show that word length is not an effective authorial discriminator but he did find similarities between Shakespeare and Marlowe which, to this day, are still being investigated albeit with vastly more sophisticated techniques!” (Holmes, 1998: 111).

²⁵ Lutoslawski (1863-1954) fue un hombre multidisciplinar y polifacético: filósofo, lingüista, místico – a él se debe también la introducción del término “metapsíquica”, para referirse a la “ciencia psíquica” – desarrolló incluso una entusiasta actividad política y fundó el partido Philaretos (Pawlowski y Pacewicz, 2004).

McMenamin menciona, en *Forensic Linguistics* (2002), los esfuerzos de D. W. Reed y G. R. Pickford, en la primera mitad del siglo XX, por dirigir la atención de la investigación lingüística hacia el fenómeno de la variación, esfuerzos que no dieron fruto, pues los desarrollos teóricos propuestos en la época preferían concentrarse en el lenguaje de “un hablante ideal, que existía tan solo como una imagen mental sin contacto con la realidad lingüística y social” (McMenamin, 2002: 137).

En 1964, el estudio de Frederick Mosteller y David Wallace sobre la autoría de los *Federalist Papers*, estableció un punto de inflexión²⁶. Mosteller y Wallace aplicaron por primera vez métodos estadísticos bayesianos combinando información de diversos rasgos textuales, lo que supuso un planteamiento por completo diferente a la hora de abordar el análisis de los rasgos de estilo y otorgó a la estilometría un prestigio que ha conservado desde entonces. Otros estudios, en cambio, resultaron más controvertidos, como los realizados sobre el *Libro del Mormón*, cuyos resultados dieron lugar a nuevos análisis, réplicas y argumentaciones que, si bien no han llegado a zanjar por completo la cuestión, sí han generado un intenso debate sobre los métodos y técnicas empleados. Sobre ambos estudios tendremos ocasión de extendernos más adelante.

En el terreno penal, el análisis estadístico de rasgos lingüísticos ha contribuido a resolver algunos casos de gran atención mediática, como el del terrorista norteamericano Theodore Kaczynski, que recibió el apodo de *Unabomber* y que entre los años 1978 y 1995 se dedicó a enviar artefactos explosivos por correo postal. El estudio que el FBI realizó del manuscrito que el autor remitió a seis publicaciones de ámbito nacional constituyó una prueba

²⁶ Fue a partir de entonces que este tipo de estudios de atribución de autoría “comenzaron a denominarse no-tradicionales, en oposición a aquellos que hasta entonces se habían apoyado en la pericia y conocimientos de investigadores humanos” (Stamatatos, 2009: 538).

decisiva que contribuyó a su condena²⁷, después de que fuese identificado gracias a cierta expresión que incluyó en el mismo y que su propio hermano reconoció como característica del sospechoso (Coulthard, 2004).

Las técnicas estilométricas están diversificando su aplicación al margen de la atribución de autoría literaria y nuevas líneas de investigación surgen a medida que se desarrolla la tecnología de la que se sirven. El análisis de texto para determinar posibles casos de plagio, la atribución de autoría en mensajes e-mail, posts en blogs y foros sociales o el estudio de códigos fuente de programación, son algunas de estas aplicaciones, en muchos casos directamente relacionadas con investigaciones policiales o judiciales (Stamatatos, 2009). Los estudiosos de la teoría literaria están comenzando a explorar las posibilidades que los estudios estilométricos ofrecen para conocer aspectos relacionados con los géneros, las técnicas de escritura o la idiosincrasia social y cultural de los autores. Los análisis de mercado y la ciencia social son también terrenos donde puede aportar información de gran utilidad. No es de extrañar, pues en el momento de escribir estas líneas, el fenómeno *big data* se encuentra en plena eclosión²⁸. Los recientes avances en

²⁷ En la cabaña en la que el detenido residía, en los bosques de Montana, el FBI encontró un documento de 300 palabras escrito varios años antes y del que los investigadores extrajeron una serie de doce rasgos lingüísticos coincidentes con el manuscrito enviado a los periódicos. La acusación presentó el informe policial como una evidencia de la autoría común de ambos textos. Por su parte la defensa contrató los servicios de un lingüista que argumentó que los ítems señalados por la policía no tenían especial significancia, dado que al ser compartidos por toda la población hablante del mismo idioma, perdían su capacidad de señalar al acusado como el autor del texto. El FBI contra-argumentó presentando un nuevo estudio en el que mostraban que, de un total de 3 millones de documentos tomados de internet y que contenían al menos una de las palabras seleccionadas por la policía en el manuscrito encontrado en la cabaña, tan solo 69 contenían las doce y todos ellos eran copias o versiones del documento remitido por el terrorista a la prensa. (Coulthard, 2004). El lingüista y profesor del Vassar College de Nueva York Donald Wayne Foster intervino como perito en este caso y, tras analizar los textos, concluyó que la evidencia en contra del acusado era irrefutable.

²⁸ Estas masas tan enormes de datos – que con frecuencia alcanzan varios pentabytes – superan la capacidad de almacenamiento y procesamiento de los sistemas informáticos habituales, lo que ha propiciado la aparición de empresas dedicadas a la gestión del *big data* estadístico o *data mining*. En estos casos, la clave para un aprovechamiento eficaz de toda esta información radica en formular las preguntas adecuadas. A estas alturas es bien conocido que uno de los factores que influyeron de forma decisiva en la reelección de Barack Obama fue el uso que su equipo de asesores, dirigido por Rayid Ghani, dió a la información que las redes sociales proporcionaban acerca de los problemas concretos que preocupaban a grupos de electores determinados, lo que permitió al candidato dirigirse a ellos de forma específica (Executive Office of the President, 2012; Bosmol Social Media News, 2013).

tecnologías de la información y comunicación, están facilitando el acceso a cantidades ingentes de datos estadísticos, que una vez analizados revelan aspectos que, hasta el momento, habían pasado inadvertidos en terrenos tan dispares como la biología, la meteorología, la medicina y por supuesto la economía, desde los mercados de valores a la gestión logística.

“Stylometric analysis is important to marketers, analysts and social scientists because it provides demographic data directly from raw text. There has been growing interest in applying stylometry to the content generated by users of Internet applications, e.g., detecting author ethnicity in social media, or whatever someone is writing deceptive online reviews.” (Bergsma et al., 2012: 327-337).

Algunos de los más recientes trabajos van enfocados, por ejemplo, a la determinación del género (masculino/femenino) del autor, si el texto está escrito o no en su L1 o si se trata de un texto destinado a una conferencia o tiene una finalidad didáctica e incluso una “rama” nueva de la estilometría está gestándose en los últimos tiempos, destinada a desarrollar sistemas que garanticen el anonimato de los autores de un texto²⁹ (Brennan, Afroz y Greenstadt, 2012; Kacmarcik y Gamon, 2006). En nuestro país, el profesor Paolo Rosso, desde la Universidad Politécnica de Valencia, donde dirige el Laboratorio de Ingeniería de Lenguaje Natural, viene realizando desde hace años un trabajo notable en el campo de la detección de plagio y la utilización del lenguaje figurativo en los medios de comunicación (Reyes y Rosso, 2014; Turell y Rosso, 2014; Bensalem, Rosso y Chilhi, 2014). Son solo algunas muestras de la versatilidad que ofrecen los resultados de los análisis estilométricos, pero que revelan una enorme diversidad en su aplicación práctica.

²⁹ Trabajos que los autores justifican desde el punto de vista ético aludiendo a la necesidad que los individuos tienen en ocasiones de denunciar públicamente abusos corporativos o estatales. Si la posibilidad de anonimato desapareciese por completo, aseguran, “el temor a represalias impediría que estas denuncias saliesen a la luz” (Kacmarcik y Gamon, 2009: 444).

2.2.2. Métodos de análisis

Así pues, a la hora de abordar la tarea de la atribución de un texto anónimo o dubitado mediante técnicas estilométricas es preciso decidir cuál de entre los diversos métodos, desarrollados a lo largo de las últimas décadas, es el más idóneo, de acuerdo a las circunstancias particulares del texto bajo escrutinio. A estas alturas, nos parece que no es preciso insistir en la gran diversidad de sistemas que han sido propuestos y en el desacuerdo que divide la opinión de los especialistas en la materia a este respecto. Esta falta de consenso es precisamente la queja más habitual entre los lingüistas forenses, sin que por ello parezca que las posturas se acerquen en lo más mínimo. Hace quince años, Joseph Rudman en “The State of Authorship Attribution Studies: Some Problems and Solutions” hacía un llamamiento a la comunidad investigadora para que aunasen esfuerzos en pro de la muy necesaria tarea de unificar conceptos, métodos de análisis y sistemas de evaluación de los resultados.

“The results of most non-traditional authorship attribution studies are not universally accepted as definitive. One major indication that there are problems in any field is when there is no consensus on results, no consensus as to accepted or correct methodology, and no consensus as to accepted or correct techniques.” (Rudman, 1998: 352).

Desde entonces, la situación no parece haber mejorado en absoluto y la disciplina sigue, en conjunto, fragmentada por una división de opiniones planteamientos y teorías divergentes, como lamenta Shlomo Argamon:

“The field of automated (non-traditional) authorship attribution is (not to put a fine point on it) a mess. While many significant accomplishments have been achieved, the field is highly fragmented, with little or no general theory or deep understanding of the strengths and weaknesses of different methods. Indeed, there is very little

agreement, if any, on standard evaluation methods, so it is nearly impossible to really measure progress in the field as a whole.” (Argamon, 2012: 95).

Así, la riqueza que toda diversidad lleva consigo resulta en este caso atenuada por la descoordinación de los esfuerzos investigadores y, sin que ello sirva de excusa, nos atrevemos a sugerir que tal vez esta falta de consenso sea en buena parte debida a la velocidad con que avanza la tecnología informática; apenas un programa ha sido sometido a prueba y puesto en práctica, cuando ya otro más potente ofrece nuevas posibilidades de análisis. En estas condiciones, en las que resulta tan difícil mantenerse al día de las herramientas tecnológicas disponibles, no es de extrañar que las líneas de investigación se dispersen y sean pocos los lingüistas que disponen del tiempo suficiente como para informarse de los trabajos más recientes y, menos aún, verificar sus resultados.

Las preferencias sobre el rasgo más relevante a la hora de analizar estilométricamente un texto escrito, por ejemplo, oscilan desde la longitud media de las palabras, hasta la riqueza de vocabulario, pasando por la frecuencia de aparición de ciertos caracteres, la combinación de palabras o la presencia de determinadas construcciones sintácticas. Joseph Rudman calcula que hasta el momento han sido propuestas unas mil mediciones diferentes (Rudman, 1998).

A pesar de todo, existen, afortunadamente, varios estudios comparativos, relativamente recientes, que revisan los métodos más empleados y dan cuenta de la efectividad de cada uno de ellos (Grieve, 2007; Juola, 2008; Koppel et al., 2009; Stamatatos, 2009; Luyckx, 2010, Savoy, 2012) y que nos serán de gran ayuda en la elección del que nosotros aplicaremos en este caso.

2.2.2.1. La extensión del texto

Tal vez el primer aspecto que deba considerarse a la hora de abordar la tarea de atribución de autoría sea la de determinar el tipo de texto de que se trata y en este sentido la extensión resulta crucial. Cuanto más breve sea un texto, menos marcadores de estilo aparecerán en el mismo y por tanto más difícil será elaborar un “perfil estilístico” de su autor, hasta el punto de convertirse en una tarea imposible si el fragmento en cuestión es demasiado reducido (Olsson, 2008). Esta es una dificultad frecuente en investigación policial, donde las notas de suicidio, cartas de amenaza o peticiones de rescate rara vez superan las doscientas palabras y con frecuencia tienen menos de cien (Coulthard, 2005a). No hay, por el momento, acuerdo unánime al respecto de cuál es la extensión mínima que permite extraer de un texto los rasgos característicos del estilo (Stamatatos, 2009), aunque se han realizado trabajos con distintos tamaños; Koppel et al., (2007), por ejemplo, emplean bloques de 500 palabras y Hirst y Feiguina (2007) han comparado los resultados obtenidos a partir de bloques de 200, 500 y 1.000 palabras, concluyendo que la precisión desciende acusadamente a medida que lo hace el tamaño de los textos. El tráfico de *e-mails* y los *posts* en blogs y foros sociales, en concreto, suponen un desafío para los lingüistas por su escasa extensión, si bien las peculiaridades idiomáticas y ortográficas y los errores gramaticales, tanto intencionados como involuntarios, se multiplican y Luyckx y Daelemans (2010) han constatado que el uso de n-gramas como marcadores ofrecen buenos resultados incluso en textos breves.

A pesar de todo, a medida que crece el número de delitos cometidos a través de internet, también lo hace el interés por desarrollar técnicas que contribuyan a la localización e identificación de los delincuentes. Los estudios estilométricos se revelan como una eficaz herramienta que puede contribuir a paliar el tradicional anonimato que protege a quienes quebrantan la ley utilizando las redes sociales. Por otro lado, no son pocas las compañías

privadas interesadas en establecer perfiles de edad, sexo y preferencias de los usuarios de internet, con fines comerciales. El análisis de los mensajes de correo electrónico, *posts* o conversaciones en *chats* sociales, sin embargo, presentan unas características propias que requieren un tratamiento específico. A la hora de estandarizar los elementos que van a ser analizados y a servir de corpus de entrenamiento para los programas empleados, resulta imprescindible, por ejemplo, eliminar las marcas *html* y sobre todo aquellas intervenciones procedentes de *chatbots* – pequeños programas que simulan conversaciones humanas, generalmente con fines comerciales – y que pueden distorsionar los resultados finales. De hecho, el estudio de los rasgos que diferencian a un autor humano de otro artificial se apoya sobre todo en estudios estilométricos, por lo que la pugna entre quienes se esfuerzan por desarrollar programas que imiten de forma realista la producción lingüística de personas reales y quienes intentan desenmascarar a estos interlocutores artificiales está impulsando la investigación en los campos de la pragmática, análisis del discurso y gramática generativa.

La cuestión de cuántos autores han intervenido en la elaboración del texto es también de gran importancia, aunque rara vez suele ser tomada en cuenta. Juola (2008) recuerda que, en teoría, solo los textos de los que se tiene la certeza de que han sido producidos por un único autor deberían ser analizados. En la práctica añade, esto rara vez se cumple, pues, cuando se trata de textos literarios, es frecuente que el manuscrito original haya sufrido la intervención de editores y compositores, que en mayor o menor medida son responsables de su “aspecto final”. Las citas o reproducciones textuales de otros autores, por ejemplo, deberían ser eliminadas para que no contaminen los resultados. Stefanova Spassova (2009), citando a Love (2002), recuerda la clasificación que este hace de posibles autorías y en la que distingue entre la *autoría ejecutiva*, que es aquella debida a un solo autor; *colaborativa*, en la que otras personas han revisado y corregido el original; *autoría de revisión*, si el propio autor ejecutivo colabora en la revisión; *declarativa*, donde un

individuo hace suyo el texto que le han preparado otros, como sucede en el caso de los discursos políticos y *precursora*, cuando un texto se nutre de otros anteriores, en forma de citas o reformulaciones.

Además, los procesadores de texto actuales, con sus correspondientes correctores, son las herramientas con las que se elaboran la mayoría de los escritos hoy en día, que se benefician de la supresión de erratas tipográficas e incorrecciones gramaticales, pero que complican la tarea del lingüista forense. Algunos autores, como Tim Grant (2008), advierten que obviar todos estos factores compromete la calidad de los resultados finales de un estudio de atribución de autoría. En la práctica todas estas cuestiones elevan las exigencias de “autenticidad” de un texto hasta niveles difíciles de alcanzar, por lo que un cierto grado de compromiso se hace necesario.

2.2.2.2. El número de autores

El número de autores candidatos es otra cuestión de gran importancia, pues cuanto mayor sea dicho número, menor será el porcentaje de aciertos o identificaciones positivas. Sin embargo, y como veremos, este porcentaje también aumenta de forma significativa en la dirección inversa, hasta rozar la certeza absoluta, cuando el panel de posibles autores se reduce.

Así pues, el peor de los escenarios al que podemos enfrentarnos es aquel en que no tenemos ningún autor candidato, lo que equivale a decir que el panel de “sospechosos” está formado por el total de la población, lo que podríamos llamar “panel abierto”. Por desgracia, y a diferencia de lo que sucede con las huellas dactilares o el ADN, no existe por el momento – y es muy probable que nunca llegue a existir (Coulthard, 2005b) – un registro con muestras indubitadas y asignadas cada una de ellas a una persona en concreto. Esta situación no es infrecuente en las investigaciones policiales, donde el

análisis de notas anónimas que deben ser verificadas se da con una cierta frecuencia³⁰. Pero aún entonces el análisis estilométrico puede revelar aspectos como el sexo del autor (Koppel et al., 2002), su edad (Burger y Henderson, 2006), su lengua materna (Koppel et al., 2006) o posibles alteraciones psicológicas (Pennebaker y King, 1999), por ejemplo, datos que permitirán elaborar un perfil (“authorship profiling”) personal que contribuya a su identificación. Además, el análisis estilométrico de textos publicados en las redes sociales está demostrando ser una eficaz herramienta para luchar contra delitos como el acoso o la pedofilia (Bogdanova, Rosso y Solorio, 2014).

La información procedente del cribado estadístico de los textos proporciona información de gran interés sobre los autores y, si bien los índices de fiabilidad son moderados, el sistema parece, como mínimo, prometedor. No cabe duda de que a medida que los procesos se perfeccionen, como resultado de sucesivos estudios empíricos, los perfiles generados serán cada vez más detallados y precisos. Entre tanto, merece la pena destacar el ensayo firmado por Argamon, Koppel, Pennebaker y Schler (2009), “Automatically Profiling the Author of an Anonymous Text”, en el que recogen los resultados de un estudio poblacional de rasgos personales a partir de textos anónimos, pues sus conclusiones son una buena muestra de la valiosa información que es posible obtener por este método y que ha servido de referencia a trabajos y estudios posteriores.

³⁰ El tristemente caso de Jack el Destripador ocurrido en Londres, entre los meses de agosto y noviembre de 1888, es un buen ejemplo. Los investigadores se encontraron con una situación especialmente enrevesada, pues tuvieron que analizar cientos de misivas de individuos en las que los autores aseguraban ser el asesino. De todas ellas, la policía centró sus indagaciones en tres, las conocidas como “Dear Boss” – en la cual el autor firmaba con el apodo que a partir de entonces se haría famoso – “Saucy Jacky” y “From Hell”. Todas ellas fueron examinadas cuidadosamente en busca de indicios que señalasen la identidad del asesino, centrándose sobre todo en el tipo de escritura con que estaban elaboradas, pero también en ciertas marcas textuales, como errores gramaticales que se sospechaba podrían haber sido incluidos de forma intencionada por el autor para disimular su verdadero nivel cultural (Evans y Skinner, 2005).

En lo relativo al género del escritor, las muestras recogidas de un total de 19.320 autores diferentes – a partir de *posts* colgados en *blogs*, que acumulan un promedio de 7.250 palabras por autor – muestran que los rasgos relacionados con el contenido resultan más reveladores que los rasgos sintácticos.

Como puede verse en la Tabla ilustrativa número 1, los rasgos de estilo que resultan de mayor utilidad a la hora de determinar el género son los determinantes y las preposiciones, para los autores de género masculino y los pronombres para los autores del género femenino. En cuanto al contenido semántico, son los términos relacionados con la tecnología en el primer caso y con la vida personal o social en el segundo. Los autores del estudio se cuidan de recordar que en lo que hace referencia a la utilidad de los aspectos semánticos como marcadores de género, el contexto es de gran importancia, por lo que textos extraídos de fuentes diferentes a los blogs, como es este caso, precisarían marcadores semánticos distintos.

Tabla 1

Class	Style Features	Content Features
Female	personal pronoun, I, me, him, my	cute, love, boyfriend, mom, feel
Male	determiner, the, of, preposition-matter, as	system, software, game, based, site
Teens	im, so, thats, dont, cant	haha, school, lol, wanna, bored
Twenties	preposition, determiner, of, the, in	apartment, office, work, job, bar
Thirties+	preposition, determiner, of, the, in	years, wife, husband, daughter, children
Bulgarian	conjunction-extension, pronoun-interactant, however, pronoun-conscious, and	bulgaria, university, imagination, bulgarian, theoretical
Czech	Personal pronoun, usually, did, not, very	czech, republic, able, care, started
French	indeed, conjunction-elaboration, will, auxverb-future, auxverb-probability	identity, europe, european, nation, gap

Russian	can't, i, can, over, every	russia, russian, crimes, moscow, crime
Spanish	determiner-specific, this, going_to, because, although	spain, restoration, comedy, related, hardcastle
Neurotic	myself, subject pronoun, reflexive pronoun, preposition-behalf, pronoun-speaker	put, feel, worry, says, hurt
Non-neurotic	little, aux verbs-obligation, nonspecific determiner, up, preposition-agent	reading, next, cool, tired, bed

Los análisis destinados a determinar la edad del autor de un texto anónimo y que hacen uso combinado de rasgos de estilo y contenido alcanzan una precisión de hasta un 77,7 %, siendo los segundos los más reveladores y predecibles en sí mismos: cuestiones relativas a la escuela y los estados de ánimo para los adolescentes, relativas al trabajo y las relaciones sociales para los veinteañeros y relativas a la vida familiar para los de treinta y tantos.

A la hora de establecer los rasgos indicadores de la lengua nativa, Argamon et al., (2009) hicieron uso del *International Corpus of Learner English*. Todos los autores de los textos de muestra y entrenamiento eran estudiantes universitarios y a la vez estudiantes de inglés como L2, con edades y nivel lingüístico similar. En este estudio en concreto, se tuvo en cuenta a participantes originarios de cinco países: España, Rusia, República Checa, Bulgaria y Francia. De nuevo y como puede verse en la Tabla 2 de resultados, son los rasgos relacionados con el contenido los que proporcionan más información; de hecho, al combinar contenido y rasgos de estilo, estos últimos lastran el resultado final.

Tabla 2

	Baseline	Style	Content	Style+Content
Gender (2 classes)	50.0	72.0	75.1	76.1
Age (3 classes)	42.7	66.9	75.5	77.7
Language (5 classes)	20.0	65.1	82.3	79.3
Neuroticism (2classes)	50.0	65.7	53.0	63.1

Con respecto al contenido, los investigadores (Argamon et al., 2009) observaron que los hablantes nativos de ciertas lenguas utilizaban ciertas palabras más que otras y una revisión somera de los resultados llama nuestra atención hacia ciertos datos: en los estudiantes de origen español, ruso, checo y búlgaro la primera palabra que aparece es la que designa su país natal – *Spain, Russia, Czech Republic, Bulgaria*, sin embargo, en el caso de los de origen francés, no aparece el nombre de su país, pero en cambio sí lo hacen dos términos que no están incluidos entre los marcadores semánticos de los de las otras nacionalidades: *europe* y *european* (Tabla 1). No entraremos a valorar aquí el significado que este dato pueda tener al respecto de la imagen que de su propia identidad nacional tienen unos y otros, pero nos parece que serviría para abrir una interesante línea de investigación.

En el apartado de personalidad, los sujetos cuyos textos fueron utilizados como muestras eran estudiantes de psicología de la Universidad de Texas. Todos ellos recibieron indicaciones para que escribiesen durante veinte minutos lo primero que les viniese a la mente, en un ejercicio de *stream of consciousness*. Todos, además, habían completado previamente un cuestionario estándar de determinación de personalidad (John, 1990). Sin embargo, Argamon et al. se centraron en único rasgo: la neurosis – enfermedad del sistema nervioso que se caracteriza por la inestabilidad emocional – y en este caso los rasgos de estilo fueron los más reveladores, mientras que los de contenido semántico resultaron prácticamente irrelevantes. A pesar de que el porcentaje de reconocimientos positivos de

rasgos neuróticos a través del análisis estilométrico de los textos son modestos, apenas un 65,7%, los autores del estudio lo ponen en contexto al recordar que los intentos de detectar rasgos neuróticos de familiares o conocidos de los sujetos, a partir del trato cotidiano, tan solo alcanza el 69%³¹.

En conjunto, y a pesar de que se trata de un estudio reducido, en relación al número de aspectos personales de los autores que podrían establecerse como objetivo, sirve para hacerse una idea de las posibilidades que el sistema ofrece a los especialistas de las distintas ramas de las ciencias humanísticas.

Por supuesto, en estas situaciones y para que el lingüista forense pueda extraer información fundada y de utilidad de los textos anónimos analizados, es preciso disponer de estudios poblacionales extensos que recojan las variaciones lingüísticas en las distintas comunidades del habla y en las que intervienen factores geográficos, de género y socio-culturales muy diversos³².

A pesar de que la situación de “panel de candidatos abierto” es muy frecuente, es preciso decir que antes o después termina por convertirse en lo que hemos denominado “**panel de candidatos cerrado**”. En esta situación, el total de autores candidatos suele estar formado por un número mucho más manejable, que puede oscilar entre dos y veinte, a lo sumo, en la mayoría de

³¹ Los progresos en este terreno deben ser, a pesar de todo, cuidadosamente evaluados. En un examen psicolingüístico de la novela *Primary Colours*, el reputado, a la vez que controvertido, lingüista de la Universidad de Vassar, Donald W. Foster, dictaminaba que el autor era “blanco, de mediana edad, ambivalente hacia las mujeres; (...) alguien a quien le gustaría educar a personas de color al respecto de lo que más les conviene...” (Foster 2001: 62-63). Un dictamen que no fue bien recibido, a pesar de que Foster identificó correctamente a Joe Klein como el autor de esta novela, publicada en un primer momento en forma anónima.

³² Este tipo de estudios requieren una gran inversión, en términos tanto económicos, como de tiempo y esfuerzo, por lo que generalmente precisan del respaldo institucional. El caso de Carol E. Chaski en un buen ejemplo, pues desde hace varios años lleva a cabo un muestreo de textos procedentes de distintos grupos sociales que el Departamento de Justicia de los Estados Unidos contribuye a financiar (Chaski, web).

los casos. El proceso que lleva a pasar de una a otra situación suele estar condicionado por circunstancias extralingüísticas: ciertos candidatos, la inmensa mayoría a decir verdad, son excluidos simplemente porque no tuvieron la oportunidad de elaborar el texto en cuestión. La carencia de medios materiales o coincidencias oportunas relativas al momento y al lugar, hacen que los investigadores descarten a una buena parte de los sospechosos “teóricamente posibles” y pasen a considerar un grupo mucho más reducido de “teóricamente probables”. Los investigadores, por otro lado, deben esforzarse por que el número de candidatos que integran este último grupo, el “**panel cerrado**”, sea lo más bajo posible, pues, como decíamos más arriba, la precisión de los resultados del análisis estilométrico aumenta en proporción inversa. Así, por ejemplo, en el estudio comparativo de Jack Grieve, “Quantitative authorship attribution: An evaluation of techniques” (2007), la precisión de un perfil basado en frecuencia de aparición de ciertas palabras – con un mínimo de diez textos de muestra por autor – oscila entre un 90% para el caso en que sean 2 los candidatos y 50% en el caso de que los candidatos sean 20.

La situación de *panel cerrado* tiene, además, sus propias peculiaridades, la primera de las cuales, y tal vez la más importante, es que al elaborarlo deberíamos tener la certeza absoluta de que el autor del texto dubitado se encuentra incluido en el mismo (Grant, 2008). Si esto no es así, todo el estudio se viene abajo, pues por mucho que uno de los candidatos con los que estamos trabajando muestre rasgos que lo señalen como el más probable, nunca podríamos tener la seguridad de que fuera del panel no se encuentre el autor real del texto. Dado que, como hemos dicho antes, las circunstancias que excluyen a unos e incluyen a otros en el *panel de candidatos cerrado* son extralingüísticas, quienes llevan a cabo estas indagaciones deben extremar la precaución para asegurarse de que ningún posible candidato quede fuera, mientras que al mismo tiempo intentan que el número de los que entran no resulte excesivo.

Una vez hecho esto, y con un número reducido de autores candidatos, el proceso a seguir pasa por obtener muestras de texto indubitadas de cada uno de ellos, cuanto más extensas, mejor. A partir de ahí, el planteamiento es sencillo, al menos en teoría: se trata de establecer cuál de los textos indubitados guarda una mayor similitud con el texto dubitado y atribuirle al autor del primero la autoría del segundo (Chaski, 2007; Grant, 2007; Eagleson, 1994). Para ello se extraen los marcadores de estilo que se consideren oportunos – y cuya selección varía de unos estudios a otros – y a partir de ahí se lleva a cabo la comparación.

“Once it has been established what texts are to be compared, the stylistic analyst has to determine whether the variance within a text is larger or smaller than between texts i.e. the analyst has to show that what is being measured varies more between works of two authors than between different works of the same author.” (Holmes, 1985: 329).

A la hora de extraer los marcadores de estilo que van a servir para elaborar el *perfil estilístico* de cada autor, es importante disponer de una cantidad suficiente textos indubitados con el fin de poder determinar con precisión qué marcadores se mantienen constantes, tanto a través de variaciones temporales, como genéricas o de contenido. Efstathios Stamatatos (2009) propone dos sistemas para el tratamiento de los textos de cada autor: o bien reunir todas las muestras de cada autor y formar un único texto por cada candidato del cual se establecerá el perfil de su estilo, o bien conservar las diferentes muestras de cada autor y tratarlas de forma individual elaborando un perfil por cada una de ellas. En el primero de los casos las diferencias entre las distintas muestras de un solo autor se desechan y Stamatatos advierte que el perfil obtenido del texto en conjunto puede diferir considerablemente del que se obtendría a partir de cada una de las muestras por separado. En la

actualidad es el segundo sistema por el que se inclinan la mayoría de los analistas.

La comparación entre los perfiles elaborados a partir de los textos dubitados e indubitados se lleva a cabo por medio de programas estadísticos que señalan cuál es el candidato que se aproxima más al texto cuestionado y a este es a quien se le atribuye la autoría. Algunos autores sin embargo, no comparten esta terminología, como es el caso de Tim Grant (2008) y de John Olsson:

“When using the term 'authorship attribution' in the context of any type of text, short or long, we cannot use it in a technical sense. This is why I suggest that the most appropriate technical term for authorship testing is authorship comparison. Some linguists use authorship identification. My reason for disliking this term is that, again, if we have two possible authors as candidates, we are not really undertaking an identification exercise. To say that our writer is more likely to be x or more likely to be y is not the same as saying that our writer is x or y. We are not actually identifying x or y. We are proposing a greater likelihood of x than y, or the other way around.” (Olson, 2008: 44-45).

Olsson pone de manifiesto que el resultado de los análisis estilométricos informa de la proximidad o similitud entre estilos, pero no hay por el momento forma de asegurar de manera terminante que ambos están escritos por el mismo autor, incluso aunque nos encontremos ante una situación de *panel de candidatos cerrado*. Por esta misma razón, McMenamin (2002) propone una escala de probabilidades a la hora de proporcionar informes forenses de autoría (Tabla 3).

Tabla 3

Did the suspect write it?	Scale	Conclusion
YES	9	Identification
	8	Highly probable – did write
	7	Probable – did write
	6	Indications – did write
INCONCLUSIVE	5	No conclusions
NO	4	Indications – did not write
	3	Probable – did not write
	2	Highly probable – did not write
	1	Elimination

Se trata de una “escala de probabilidad de autoría” muy similar a la “escala de opinión” de los expertos que llevan a cabo el análisis y que Coulthard y Johnson (2012) sugieren para su uso en informes de lingüística forense ante los tribunales, dado que la atribución, en sentido estricto, es imposible, tal y como recuerda Olsson.

Merece la pena que nos detengamos durante unos instantes a reflexionar en el procedimiento mencionado, pues es muy revelador de la realidad con la que se trabaja la atribución de autoría por medios estadísticos. Recordemos que en principio, debemos tener elaborado un *perfil estilométrico* de cada uno de los textos candidatos y otro del texto dubitado. El objeto de este perfil es atenuar las variaciones intra-autor que puedan producirse por circunstancias contextuales y puede sernos de ayuda visualizar dichos *perfiles* como si de retratos fotográficos se tratase, como si hiciésemos un retrato fotográfico de una persona que no está quieta. Al fin y al cabo, el estilo de escritura de un autor está siempre *moviéndose*, variando en mayor o menor medida. Ahora bien, si estos retratos fotográficos – *perfiles estilométricos* – fuesen nítidos, bastaría superponer uno sobre otro, los retratos candidatos sobre el retrato dubitado, y cuando dos de ellos coincidiesen, podríamos afirmar que estaríamos ante la misma persona. Sin embargo y por desgracia, la atribución de textos no funciona así, ya que los *perfiles* – los retratos fotográficos – no son nítidos, con lo cual al superponer unos sobre otros, solo

cabe establecer *parecidos razonables* entre aquellos cuyos rasgos más se aproximen.

Esto significa que, en la práctica, es materialmente imposible establecer un perfil del estilo de escritura o habla nítido y preciso. A lo más que podemos aspirar es a reducir el grado de imprecisión, pero siempre será un retrato *un poco borroso*³³. Esto puede parecer frustrante, pero a modo de consuelo vale la pena recordar que es la forma en que la estadística contempla el mundo. Nada es fijo o estable; a lo sumo hay una probabilidad muy elevada de que lo sea. Se trata, además, de una descripción de la realidad cada vez más extendida y empleada por todo tipo de ciencias y disciplinas, de la medicina a la gestión de mercados y de la física cuántica a la toma de decisiones militares. Si queremos un ejemplo práctico e inmediato, solo tenemos que preguntarle a un meteorólogo el tiempo que va a hacer al día siguiente; su respuesta vendrá siempre en tantos por ciento de probabilidades de lluvia o de buen tiempo.

La última posibilidad, en cuanto a número de candidatos, la constituye el caso en que tan solo hay uno y la tarea del lingüista consiste en confirmarle o rechazarle como autor un texto anónimo, o que se atribuye a otro escritor, como sucede en el presente trabajo. Esta situación, que podríamos llamar de “**panel de candidato único**” y que en inglés acostumbra a denominarse “authorship verification”, presenta también sus propias peculiaridades. No existe aquí la necesidad de asegurar la inclusión del autor en el panel de candidatos, por cuanto el objeto del estudio consiste tan solo en determinar si el escritor analizado es o no el autor del texto dubitado. Para Koppel et al. (2009), esta situación constituye un reto sobre el que se ha estudiado muy poco y es considerablemente más complicado que el de “panel cerrado”, pues, en efecto, el método *de aproximación* que se utilizaba cuando se

³³ Si se tratase de un punto en el espacio, estaría siempre moviéndose y nos sería imposible superponer el punto del texto indubitado, que estaría también moviéndose. No tendríamos más remedio que atribuir la autoría del *punto dubitado* al *punto indubitado* que estuviese más próximo.

disponía de un número determinado de candidatos, entre los que se tenía la certeza de que se encontraba el autor, ya no es de utilidad. ¿Cómo de cerca deberían estar los dos textos, el dubitado y el indubitado, para poder afirmar que han sido escritos por el mismo autor? ¿Cuál será el criterio de *atribución*, si ya ha quedado claro que no es posible obtener un perfil nítido e invariable?

“The situation in which we suspect that a given author may have written some text but do not have an exhaustive list of alternative candidates is a common one. The problem is complicated by the fact that a single author may vary his or her style from text to text or may unconsciously drift stylistically over time, not to mention the possibility of conscious deception. Thus we must learn to somehow distinguish between relatively shallow differences that reflect conscious or unconscious changes in an author’s style and deeper differences that reflect styles of different authors.” (Koppel et al., 2009: 14).

En realidad, el problema de la autoría del “candidato único” podemos sintetizarlo como sigue: Incluso aceptando la existencia del idiolecto, la personal e idiosincrásica variedad lingüística de que cada hablante hace uso, y la existencia de ciertos rasgos idiolectales – o marcadores de estilo – que se hayan presentes en la producción de cada hablante, y que estos rasgos – o marcadores – son introducidos en dicha producción de forma inconsciente, como consecuencia de sus particulares procesos cognitivos, ni siquiera entonces podríamos tener la certeza de que no existe otro hablante con un idiolecto idéntico o, al menos, muy similar. De hecho, podrían existir más de uno, tal vez diez, cien o mil; no lo sabemos. Y seguiremos sin saberlo mientras no se realicen estudios demográficos idiolectales que recojan muestras del habla de grupos lingüísticos lo suficientemente extensos como para inferir que los resultados pueden hacerse extensivos a toda la población.

Por el momento, ni siquiera existe una teoría que describa, explique y prediga de forma satisfactoria la producción idiolectal. William Labov y M.A.K. Halliday han dado los primeros pasos para llegar a comprender la mecánica que regula este fenómeno lingüístico, pero todavía quedan muchas preguntas sin respuesta. Los rasgos o marcadores presentan una variación inter-hablante, tanto en tipo, como en número, y algunos de ellos varían a lo largo del tiempo y no hay forma de saber qué y cuándo aparecerá en el habla de un individuo concreto.

En estas condiciones ante la pregunta: ¿Escribió el autor A el texto X o no? Solo podremos responder con certeza en el caso de que la respuesta sea negativa. Si tras haber examinado varias muestras indubitadas del autor A – o una sola lo bastante extensa – hemos encontrado una serie de marcadores que se mantienen constantes a lo largo de toda su producción, y si comprobamos que dichos marcadores están ausentes, o no coinciden en frecuencia de aparición y número, del texto X, entonces podremos afirmar que A no escribió X.

Por el contrario, incluso aunque encontrásemos esos mismos marcadores, en frecuencia de aparición y número en el texto X, no podríamos tener la certeza absoluta de que A fue su autor, pues, ¿cómo podríamos estar seguros de que no existe alguien más que escribe introduciendo esos mismos marcadores? A lo sumo podríamos afirmar que existen una probabilidad muy alta de que A sea el autor. ¿Cómo de alta? Tendríamos que llevar a cabo el estudio demográfico del que hemos hablando antes para hacernos una idea de en qué proporción y frecuencia aparecen dichos marcadores entre la población en general.

Así pues, no es de extrañar que, como aseguran Koppel et al., los estudios y bibliografía sobre el escenario del “candidato único” sean tan escasos, pues la gran mayoría se centran en aspectos técnicos y estadísticos

acerca de la forma en que los resultados se ven afectados por las variaciones en cuanto a la extensión del texto, el número de candidatos o el tipo de marcador empleado en el análisis. Es de este último aspecto del que nos ocuparemos a continuación.

2.2.2.3. Los marcadores de estilo

A la hora de plantear un estudio de autoría, tan importante como la extensión del texto o el número de autores candidatos lo es igualmente el tipo de marcador de estilo que tomaremos como referencia. Existe una abundante bibliografía que recoge los numerosos estudios que se han llevado a cabo intentando determinar cuáles pueden ser los más relevantes (Argamon, 2012, 2009; Cicres y Turell, 2012; Coulthard, 2004; Fucks, 1952; Grant, 2008; Hirst y Feiguina, 2007; Holmes, 1985, 1998; Koppel et al., 2002; Luyckx, 2010; McMenamin, 2002; Olsson, 2008 Pennebaker y King, 1999; Spassova, 2009, Stamatatos, 2009; Turell, 2008). Holmes (1985: 330) considera que estos marcadores deben ser “prominentes, estructurales, frecuentes, fácilmente cuantificables y relativamente inmunes al control consciente”. Este último aspecto, en concreto, resulta de gran importancia, pues serán tanto más identificativos, cuanto menos hayan intervenido las funciones volitivas del autor en su selección, ya que de esta manera es de esperar que aparezcan en todos sus escritos, sea cual sea el contexto. Podríamos afirmar que a mayor control del individuo sobre su producción lingüística, más difícil será la tarea del lingüista forense.

“Essentially, computational stylistics assumes that a writer has a finite number of words at his/her disposal and, at any given moment, each writer will have certain subconscious habits of using those words, depending on such things as context and literary genre and subject to the constraints of syntax. An analyst searches for features particularly

to a given writer, features of which the writer is probably unaware and which can be measured quantitatively in order to have a basis for comparison with other write.” (Holmes, 1985: 329).

En cualquier caso una primera clasificación taxonómica de los posibles marcadores no es difícil, dado que los textos escritos están constituidos por un número finito, y reducido, de elementos; la unidad básica la forman los grafemas, entre los que se encuentran las letras, que combinadas forman las palabras, las cuales a su vez dan lugar a las frases y es con estos elementos, en solitario o combinados, con los que va a trabajar la estilometría. A pesar de todo, se han propuesto diferentes aproximaciones a esta cuestión que, de una u otra manera, reordenan las mismas categorías. Así por ejemplo, Stamatatos (2009) lleva a cabo una división entre caracteres (en los que incluye además de las letras, los signos de puntuación), elementos léxicos, elementos sintácticos y elementos semánticos, lo que en esencia cubre el total de posibilidades. Jack Grieve (2007) aplica la siguiente clasificación: un *grafema* es una letra del alfabeto, un *carácter* es una unidad indivisible e incluye grafemas, dígitos, signos de puntuación y espacios en blanco; una *palabra* es una secuencia de grafemas y/o dígitos, una *frase* es una secuencia de caracteres, excluyendo signos de interrogación y exclamación, una *n-palabra* (“n-word”) es una secuencia de *n* palabras, un *n-grama* es una secuencia de *n* caracteres y un *perfil* es un conjunto de mediciones textuales.

En cualquier caso, se elija el elemento que se elija, la función del mismo, dentro del análisis cualitativo, es la *discriminación* entre los textos comparados. Veremos a continuación las estrategias con que distintos autores abordan la elección de los elementos con los que llevarán a cabo dicha discriminación.

A. Caracteres

En este caso los textos son vistos como secuencias de caracteres, cuya aparición puede ser medida de varias formas. Grieve (2007) establece cuatro medidas diferentes para los grafemas:

1. El perfil del grafema simple, consiste en la frecuencia relativa de una letra calculada dividiendo el número de veces que aparece la letra en cuestión entre el número total de letras del texto.
2. La segunda medición es similar, pero en esta ocasión se calcula a partir de la posición que ocupa una letra dentro de las palabras que integran el texto, dividiendo el número de veces que aparece esta letra en una posición concreta – la primera letra, la segunda... la última – entre el número total de palabras.
3. Lo que Grieve llama un perfil de grafema en multiposición, y acumula varias posiciones simples, como las tres primeras posiciones dentro de una palabra.
4. La cuarta medida se elabora a partir del porcentaje de palabras que contienen cada una de las letras del alfabeto, calculado el número de palabras que contienen al menos una vez una letra determinada, entre el número total de palabras.

El uso de n-gramas de caracteres ha resultado una herramienta versátil en la atribución de autoría. Un ejemplo de un n-grama de cuatro caracteres del comienzo de este párrafo sería |El_u|, |so_d|, |e_n-|, etc. Este tipo de marcadores ha sido muy utilizado, sobre todo en la forma de n-gramas de varios caracteres (Kjell, 1994a, 1994b; Kjell et al., 1995; Ledger y Merriam, 1994; Clement y Sharp, 2003; Houvardas y Stamatatos, 2006). Esta combinación ha sido empleada también, por ejemplo, para determinar la L1 del escritor (Zigdon, 2005) y Chaski (2005; 2007) la ha empleado en tareas de atribución forense. Stamatatos (2009) considera que una de las ventajas de

este sistema es su tolerancia al ruido, como en los casos en los que textos provenientes de e-mails o mensajes extraídos de internet contienen errores gramaticales o se hace en ellos un uso heterodoxo de la puntuación. En estas situaciones, el empleo de n-gramas de caracteres evita que la representación se vea afectada en exceso.

Los estudios realizados por Grieve (2007) muestran la elevada efectividad en atribución de los n-gramas de caracteres, oscilando entre un 65% de precisión para n-gramas de dos caracteres – bigramas, que el autor considera de gran utilidad en el análisis de textos periodísticos – y 40 candidatos, y un 94% de precisión para el mismo n-grama y 2 candidatos. Aquí interviene como factor determinante la longitud asignada a n , ya que cuanto mayor sea el número de caracteres, mayor será también la información léxica y contextual proporcionada, si bien la precisión descende a medida que aumenta n , como puede verse en la Tabla 4. Se trata, además, de un valor dependiente de la naturaleza de los lenguajes naturales y que se ve alterado al pasar de un idioma a otro. Así, por ejemplo, Blasco y Ruiz (2009) muestran que si bien en inglés la unidad ideal es el bigrama, en español la precisión aumenta progresivamente a medida que lo hace también el número n del n-grama. Sin embargo, Stamatatos (2009) cita los trabajos de Sanderson y Guenter (2006) que muestran que el uso de varias secuencias de n-gramas de longitud variable – hasta 4-gramas – dan un excelente resultado cuando se analizan textos cortos en inglés.

En su tesis sobre la aplicación de los bigramas y trigramas en lengua española, Stefanova Spassova concluye que su potencial discriminatorio es elevado, “sobre todo en la creación de perfiles forenses de identificación” (Spassova, 2009: 295). Su estudio se lleva a cabo tanto sobre textos literarios, como sobre artículos de opinión:

“Los resultados de las pruebas realizadas con los textos de novela han demostrado que tanto los bigramas como los trigramas discriminan con un alto porcentaje de clasificación correcta de los casos por autor (de entre 80 y 95%) y de todos los casos (88% para los bigramas y 92% para los trigramas). El potencial discriminatorio además no acusa ninguna relación de dependencia del número de textos y autores empleados en el análisis.” (298).

Resulta interesante observar que en el caso de su aplicación en textos periodísticos, cuya extensión es mucho menor y rara vez supera las 300 palabras, la efectividad discriminadora de bigramas y trigramas se mantiene, acusando leve descenso de tan solo un 1% en ambos casos.

En cuanto a los signos de puntuación, Grive (2007) mide cinco frecuencias de aparición, de las cuales las tres primeras son variaciones de una marca simple, tomada de un total de ocho diferentes (. , : ; - ? ‘ (), si bien el cálculo para cada una de ellas es diferente. Para las tres primeras la frecuencia relativa de aparición se calcula dividiendo la frecuencia de cada marca en un texto entre el número total de caracteres. En el segundo caso, el divisor es el número total de signos de puntuación; en el tercero, el número total de palabras. La cuarta medición consiste en un perfil integrado por un signo de puntuación y una letra (*grapheme*) y la quinta, un perfil compuesto por un signo de puntuación y una palabra.

B. Palabras

La utilización de las palabras, en sus distintas formas de aparición, se remonta a los primeros ejercicios de atribución de autoría de Mendelhal (1887), quien intentó emplear la longitud de las mismas como rasgo discriminatorio (Stamatatos, 2009). En sus estudios sobre textos de

Shakespeare, Bacon y Marlowe, la longitud de las palabras característica de cada uno de ellos mostraba que las obras de Christopher Marlowe coincidían con las de Shakespeare, casi tan bien como las de este último lo hacían consigo mismas (Holmes, 1985).

Más tarde, Brinegar (1963) empleó este mismo sistema para sugerir que Mark Twain no escribió “The Quintus Curtius Snodgrass Letters”, mientras que Mosteller y Wallace (1964) pusieron a prueba la validez de este sistema en su estudio sobre *The Federalist Papers*, concluyendo que la relación entre frecuencia de aparición de palabras de una longitud determinada y contexto hacía que sirviese tan solo, y en el mejor de los casos, de aproximación (Holmes, 1985). Trabajos posteriores, como el de Smith (1983), confirmaron estas conclusiones y añadían que cuando trabajos del mismo género escritos por autores contemporáneos eran comparados utilizando la longitud de las palabras, los resultados eran tan similares que parecían haber sido escritos por el mismo autor.

La longitud media de las palabras de un texto también ha sido evaluada a partir del número de sílabas. Los trabajos de Fucks (1952), y Fucks y Lauter (1965), mostraban que la frecuencia relativa de las palabras compuestas por un número determinados de sílabas señalaban distinciones de género e idiomas, más que autores específicos (Holmes, 1985). En un estudio posterior, Brainerd (1974) encontró cierta interdependencia entre palabras consecutivas con un determinado número de sílabas, pero que esta relación variaba sustancialmente de acuerdo al tipo de texto de que se tratase, narrativo, transcripción de conversaciones, etc. y que por tanto el sistema podría ser de utilidad a la hora de elaborar clasificaciones genéricas.

Sin embargo, en lo relativo al uso de las palabras como unidad de referencia discriminadora, la cuestión de su clasificación léxica podría constituir una sub-sección en sí misma. Para Koppel et al., (2009), las

palabras de contenido semántico pueden ser de ayuda a la hora de mostrar patrones de selección léxica, tal y como indican los trabajos de Martindale y McKenzie (1995), Craig (1999), Waugh et al. (2000) y Argamon et al. (2009). Sin embargo, advierten que este tipo de marcadores – cargados semánticamente – guardan una gran dependencia del contexto, por lo que sufren grandes variaciones al cambiar el tema tratado en los textos analizados.

Para Stamatatos (2009), una clasificación basada en la distribución léxica requiere una menor dimensión en comparación con una clasificación en base al tema tratado, especialmente si se utilizan palabras funcionales, aquellas que no llevan carga semántica – artículos, pronombres, conjunciones y preposiciones. Holmes (1985) considera que las palabras funcionales proporcionan una gran fiabilidad en tareas de discriminación y menciona el uso que Ellegard (1962) hace de ellas en su estudio de autoría de las *Junior Letters*, donde calcula el “*distinctiveness ratio*” con la siguiente fórmula:

$$\text{distinctiveness} = \frac{\text{relative frequency of the word in Junius}}{\text{relative frequency of the word in a million} - \text{word sample of non - Junian writings}}$$

Los estudios que coinciden en aconsejar su utilización son numerosos: Mosteller y Wallace (1964), Morton (1965), Burrows (1987), Karlgren y Cutting (1994), Merriam y Mathews (1994), Kessler et al. (1997), de Vel et al. (2001), Holmes et al. (2001), Baayen et al. (2002), Binongo, (2003), Juola y Baayen, (2003), Argamon y Levitan (2005). Actualmente está bien asentada su relevancia como marcadores de estilo en los trabajos de atribución, dado que su utilización parece determinada por procesos cognitivos ajenos a la actividad consciente del escritor y por tanto su relación con el tema tratado en el texto es nula.

Stamatatos menciona la riqueza de vocabulario como otro elemento que ha de tenerse en cuenta al tratar con marcadores léxicos. El

procedimiento es bastante simple y funciona a partir del ratio *type/token*, $V/N=R$, donde V es el tamaño de vocabulario (número de palabras diferentes) y N es el número total de palabras que aparecen en el texto, incluidas las *hapax legomena*, las palabras que aparecen una sola vez (de Vel, Anderson, Corney y Mohay, 2001). Sin embargo, la variedad léxica depende mucho de la extensión del texto y, si bien aumenta al principio, a partir de cierto punto se produce un paulatino descenso. Para Holmes, R está limitado por el tamaño del texto de muestra. Mientras que N puede aumentar sin cesar, el vocabulario empleado en conjunto está en la práctica limitado. Como posible solución para que el *type/token* sea de utilidad, propone asignar valores fijos a N , como 500 palabras, por ejemplo.

Algunos autores, como Tweedie y Baayen (1998), han intentado establecer una fórmula que proporcione uniformidad a esta distribución, utilizando K (Yule, 1944) y R (Honore, 1979), con escasos resultados, por lo que este parámetro por sí solo se suele considerar poco fiable. La fórmula de Yule (K) acostumbra a aplicarse con el fin de determinar hasta qué punto el texto analizado ofrece diversidad léxica (Blasco y Urbón, 2009: 39).

$$Yule (K) = \frac{10^4 (\sum i^2 V_i - N)}{N^2}$$

N (tokens) = N° total de palabras del texto

donde $V_i = N^\circ$ de types que ocurren i veces

Por su parte la fórmula de Honoré (H) nos indica qué posibilidades existen de que un término ya aparecido en un texto, vuelva a hacerlo más adelante.

$$\text{Honoré (H)} = \frac{100 \log N}{1 - (V_1 / V)}$$

N (tokens) = N° total de palabras del texto

donde V (types) = N° total de formas del texto

V_1 = N° types usados una sola vez

Grieve (2007) propone tres tipos de medida de frecuencia de aparición, el primero de los cuales es idéntico al mencionado más arriba. En el segundo se considera la posición relativa de una palabra dentro de la frase, por ejemplo, el primer puesto, el segundo, la última palabra de la frase, etc. El cálculo se realiza dividiendo la frecuencia de aparición de esta palabra en este lugar concreto entre el número total de frases que contienen esta palabra. La tercera medida es un perfil de multi-posición y se calcula a partir del número de veces que aparece una palabra en un grupo de posiciones concreto, por ejemplo, los cuatro primeros lugares de una frase. Grieve cita algunos de los estudios que han considerado este factor de posición: Burrows (1992), Burrows y Craig (2001), Burrows y Hassall (1988), Holmes et al. (2001), Kenny (1978), Morton (1965), Mosteller y Wallace (1964), y Tweedie et al. (1998).

La posición relativa de una palabras con respecto a otras también ha sido considerada y para medirla se han empleado vectores de n-gramas de palabras (Peng et al., 2004; Sanderson y Guenther, 2006; Coyotl-Morales, Villaseñor-Pineda, Montes y Gómez, y Rosso, 2006). No hay, sin embargo, una fórmula sencilla para determinar cuál es la combinación de palabras más significativa, a lo que hay que añadir que este sistema es más sensible al contexto de donde se extraen las muestras (Stamatatos, 2009).

Por último, vale la pena mencionar que Koppel y Schler (2003) proponen la medición de las erratas como elemento discriminante. Para ello establecieron una serie de errores ortográficos y tipográficos y una

metodología para extraer estos marcadores mediante un corrector de textos que, según Stamatatos (2009), no llega a funcionar satisfactoriamente. Se trata de un elemento que es utilizado cuando la comparación entre textos se lleva a cabo mediante expertos humanos.

C. Sintaxis

Las estructuras sintácticas vienen siendo utilizadas como elementos marcadores de estilo desde la década de 1990 (Baayen, van Halteren y Tweedie, 1996) y su aplicación está soportada por la teoría de que cada escritor acostumbra a emplear patrones sintácticos similares, de forma inconsciente. Su empleo además se relaciona con el uso de palabras funcionales, que aparecen regularmente en determinadas estructuras sintácticas. Stamatatos (2009), previene de la necesidad de utilizar herramientas “robustas y precisas” en el tratamiento de lenguajes naturales y de que este marcador depende de forma directa del idioma en que está escrito el texto.

Estas herramientas se basan, por lo general, en un corpus anotado para elaborar árboles sintácticos a partir de cada frase, en los que aparece además la clase a la que pertenece cada palabra y de dónde se generan índices de frecuencia de aparición.

De acuerdo a Stamatatos (2009), los experimentos con este tipo de marcadores han demostrado que funcionan mejor que las mediciones de variedad léxica. Stamatatos et al. (2000; 2001) proponen otro sistema para extraer información sintáctica, que en este caso requiere de una herramienta específica, un procesador de lenguaje natural, para detectar frases y cláusulas del griego moderno. La herramienta proporciona en un primer momento elementos simples y posteriormente combina dichos elementos para producir

resultados más complejos. Los investigadores insisten en los requisitos del sistema: una herramienta específica, diseñada para trabajar con un idioma concreto.

Otros tipos de herramientas pensadas para realizar funciones similares han sido propuestas por Luyckx y Daelemans (2005) y Hirst y Feiguina (2007). Koppel y Schler (2003) proponen utilizar la información sintáctica recogida a partir de los errores cometidos por el escritor, utilizando para ello un corrector de textos comercial. Sin embargo, el resultado no ha sido del todo satisfactorio.

Para Koppel et al. (2009), las mediciones sobre la frecuencia relativa de determinadas construcciones sintácticas constituyen un método rápido y fiable de procesar los textos y mejoran considerablemente los resultados basados exclusivamente en las palabras.

En cuanto al uso de la categoría gramatical de las palabras, David I. Holmes considera que proporcionan rasgos estilísticos significativos y cita a Sommers (1966):

"A more cultivated intellectual habit of thinking can increase the number of substantives used, while a more dynamic empathy and active attitude can be habitually expressed by means of an increased number of verbs. It is also possible to detect a number of idiosyncrasies in the use of prepositions, subordinations, conjunctions and articles." (Holmes, 1985: 332).

Otros, como Brainerd (1973; 1974) han constatado que, si bien las categorías gramaticales no son demasiado prácticas a la hora de señalar al autor de un texto concreto, sí que se muestran de gran ayuda en la distinción automática de géneros. Los textos procedentes de los periódicos y las

exposiciones formales, por ejemplo, tienden a emplear más artículos que las novelas, misivas y los diarios personales.

D. Frases

La longitud de las frases como posible marca de estilo de un escritor concreto fue investigada desde los primeros estudios estadísticos. Yule (1938), fue de los primeros en hacerlo, solo para concluir que no constituía un sistema fiable. Otros intentos como el de Williams (1940) no tuvieron mucho más éxito, a pesar de lo cual, a lo largo del siglo pasado se siguieron estudiando las posibilidades de la medición de longitud de frases (Wake, 1957; Morton, 1965; Sichel, 1974 y Kjetsaa, 1979).

La mayor desventaja de este tipo de marcador radica en que está sujeto, en buena medida, al control consciente por parte del escritor y dado que depende de la distribución de la puntuación, la influencia del editor del texto puede desvirtuar los resultados. Para Smith (1983), así como para Kjetsaa y Tallentire (1972), este tipo de información puede servir a lo sumo para corroborar las conclusiones obtenidas por otros medios, aunque puede ser interesante en sí misma y de utilidad para otros propósitos³⁴.

En cualquier caso, este parámetro sigue siendo empleado, a la par que la longitud de las palabras o la distribución de n-gramas, en métodos de análisis multi-variantes y Grieve (2007) propone cuatro tipos de mediciones. En la primera de ellas, se calcula la longitud media, esto es, el total del número de palabras entre el total del número de frases. La segunda mide la

³⁴ Actualmente, la longitud de las frases, junto con la de las palabras, es uno de los principales parámetros empleados en los diversos sistemas de cálculo de la legibilidad de un texto, como el Dale-Chall, el Gunning-Fog o el Flesh-Kinkaid. Por otro lado, los procesadores sintácticos constituyen la base de los buscadores de internet, como Google, y los más recientes programas de parcelado sintáctico son capaces de establecer la estructura de una frase, incluso aunque contenga palabras cuyo significado se desconoce, simplemente a partir de la posición que ocupan las palabras funcionales (Goldberg y Orwant, 2013).

distribución relativa de las frases de una, de dos, de tres palabras, etc. La tercera medición expresa la longitud media en número de caracteres, en lugar de en número de palabras. Por último, la cuarta medición es la distribución relativa de frases de entre 1 y 10 caracteres, por ejemplo, de entre 11 y 20 caracteres... Resulta interesante observar que la precisión de los análisis es muy similar tanto si se utilizan palabras para medir la longitud de las frases, como si se utilizan letras y muy baja, en cualquiera de los dos casos.

Resultados del análisis comparativo de Jack Grieve.

Grieve (2007) sometió a estudio los principales sistemas de atribución, comparando los resultados de 39 de los más empleados, algo que hasta entonces nunca se había hecho. Para ello, tomó como datos los producidos por un algoritmo básico de atribución a partir de un texto anónimo y un conjunto de textos procedentes de autores candidatos y los sometió a comparación. El algoritmo compara los valores de las medidas del texto anónimo con los valores correspondientes para cada candidato para determinar cuál es el que más se aproxima.

La Tabla 4 de resultados que se muestra a continuación, elaborada por Grieve, es una excelente referencia para determinar el grado de eficacia de cada uno de los parámetros evaluados en función del número de autores candidatos.

Tabla 4

Nº de autores/ Marcadores evaluados	40	20	10	5	4	3	2
Perfil de palabra y puntuación (límite 5)	63	72	80	87	89	92	95
2-grama perfil (límite 10)	65	72	79	86	88	91	94
3-grama perfil (límite 10)	61	72	78	85	88	91	94

4-grama perfil (límite 10)	55	64	73	83	85	89	93
Perfil de grafema y puntuación	50	60	70	81	84	87	93
Perfil de palabra (límite 5)	48	57	67	77	80	85	88
5-grama perfil (límite 10)	47	55	66	76	79	84	90
Perfil grafema multiposición (primeras 6 en palabra)	43	53	64	76	79	84	90
Perfil grafema multiposición (últimas 6 en palabra)	42	52	63	74	79	83	90
Perfil de signo de puntuación	34	46	58	72	76	80	89
6-grama perfil (límite 10)	35	45	56	68	72	78	86
Perfil de grafema en el interior de palabra	28	39	51	65	70	76	85
Perfil de grafema de posic. única (última en palabra)	27	36	49	63	68	73	84
Perfil de grafema	25	35	47	62	67	74	83
7-grafema perfil	34	42	45	59	64	69	81
Perfil de graf. Posición simple (anteúltima en palabra)	23	31	43	57	63	70	81
Perfil de graf. posición simple (primera en palabra)	20	30	41	56	62	69	80
Perfil de multi-posic. de palabra (primeros 4 en frase)	22	31	41	55	60	67	77
Perfil de longitud de palabra	18	26	39	54	60	68	79
Perfil de posición única en palabra (1ª en frase)	17	30	36	50	56	64	75
8-grama perfil	18	24	36	50	55	62	74
Perfil de colocación de 2 palabras	17	24	34	48	54	61	74
Tuldava LN	11	18	31	49	55	64	77
Perfil de longitud de frase (12 interv. 25 caract.)	12	20	31	46	53	62	74
Perfil de longitud de frase (10 interv. 5 palabras)	10	17	28	44	50	59	73
9-grama perfil (límite 2)	12	18	28	41	46	55	68
Type-token ratio	8	16	27	44	51	61	75

Herdan C	7	14	25	42	49	59	73
Guiraud R	7	13	24	41	48	58	73
Longitud media de palabra	7	12	22	39	46	55	70
Longitud media de frase (en caracteres)	6	12	22	39	45	53	70
Longitud media de frase (en palabras)	6	10	21	37	44	53	69
Yule K y Simpson D	6	10	18	33	38	49	65
6-grama perfil (límite 10)	35	45	56	68	72	78	86
Perfil de grafema dentro de palabra	28	39	51	65	70	76	85
Perfil de graf. de posic. simple (última en palabra)	27	36	49	63	68	73	84

Comparativa de los resultados de las diferentes mediciones. A la derecha se listan los marcadores utilizados y a la derecha los porcentajes de precisión obtenidos para el número indicado en la fila superior de candidatos.

A partir de los resultados obtenidos, Grieve propuso el siguiente protocolo para la resolución de casos de atribución de autoría.

1. El investigador debe confeccionar el panel de autores candidatos haciendo uso de cuanta evidencia externa o extralingüística pueda conseguir.
2. El investigador debe compilar un corpus de textos indubitados de los autores candidatos, lo más amplio y similar al texto dubitado que sea posible.
3. El investigador debe someter a prueba sobre los textos indubitados que formaban el corpus citado diversos algoritmos de atribución con el fin de establecer cuáles dan mejor resultado.
4. El investigador debe someter a prueba la combinación de aquellos algoritmos que mejor resultados hayan proporcionado.

5. Por último, de las combinaciones puestas a prueba, la que mejor funciona debe ser puesta en práctica para comparar los textos de los autores candidatos con el texto anónimo.

Como resulta evidente, la precisión de las técnicas de atribución que se apoyan en medidas estadísticas depende en gran medida del número de autores candidatos y del tamaño de los textos analizados; los indubitados y el dubitado o anónimo. Stamatatos (2009) considera que falta algún tiempo para que los análisis lingüísticos obtengan el mismo reconocimiento en los tribunales que otras técnicas forenses³⁵. Como diversos autores han observado, es necesario un mayor acuerdo en cuestión de sistemas de medición y evaluación de resultados, pero sobre todo, añade Stamatatos, necesitamos un mayor nivel de abstracción que explique las diferencias de estilo entre autores. Reclama, además y para obtener la confianza de las autoridades judiciales, técnicas de atribución más robustas, que permitan hacer frente no solo a situaciones de *panel cerrado*, sino también a las de *panel abierto* y que permitan igualmente trasladar los resultados de un género a otro; por ejemplo, que los textos de muestra procedan de los *post* colgados en un blog, pero que permitan analizar una nota de suicidio anónima.

En esencia, y como afirma Olsson, debemos recordar que la estilometría nos permite *comparar* antes que *atribuir* y que en dicha comparación estaremos hablando siempre de grados de similitud. A pesar de todo, y como también mencionábamos más arriba, esto es algo que probablemente vaya a seguir siendo así durante algún tiempo, pues la propia naturaleza de los elementos sometidos a examen no permiten obtener una certeza absoluta. La cuestión radica, por tanto, en encontrar las técnicas que

³⁵ Por desgracia para la disciplina, algunos sonados fracasos han dañado la reputación de las técnicas estilométricas. El método *cusum*, o *Qsum*, abreviatura de *cumulative sum* es un buen ejemplo. Se trata de una técnica de análisis estadístico secuencial, desarrollado por E.S. Page y aplicado a la atribución de autoría forense por Andrew Morton en varios procesos judiciales en los tribunales de Gran Bretaña. La técnica fue muy cuestionada y recibió “el golpe de gracia” cuando Morton fue invitado a una demostración televisiva en directo, donde no fue capaz de distinguir entre los escritos de un individuo convicto y el *Chief Justice of England* (Juola, 2008).

nos ayuden a incrementar el grado de probabilidad hasta niveles con los que nos sintamos razonablemente seguros, que nos proporcionen al menos la misma razonable certeza con la que salimos de casa cada día, confiando en que las posibilidades de sufrir un accidente son muy, muy pequeñas.

2.3. La Estilometría en acción: Tres casos prácticos.

De entre los muchos ejemplos de aplicación de técnicas estilométricas que pueblan la ya abundante bibliografía sobre atribución de autoría, hemos considerado de especial interés los que a continuación relatamos debido a su importancia dentro de esta disciplina. En el primero de ellos, se discute la autoría del libro canónico de cierta confesión religiosa, lo que lleva a que los resultados de los sucesivos análisis a que ha sido sometido se hayan revisado una y otra vez meticulosamente. En el segundo caso, los textos en cuestión forman parte del acervo histórico y cultural de los Estados Unidos, por lo que, al igual que en el caso anterior, los investigadores pusieron un especial cuidado en que los métodos empleados se ajustasen a protocolos científicos estrictos que garantizasen la objetividad y fiabilidad de los resultados.

El último y más reciente caso de los tres, lo constituye el relacionado con la novela *The Cuckoo's Calling*, publicada inicialmente bajo el seudónimo de Robert Galbraith, pero cuya autoría, como se supo más tarde, corresponde a la popular escritora británica J.K. Rowling. El proceso de análisis estilométrico que el reputado especialista Patrick Juola llevó a cabo sobre esta obra por encargo del diario *Sunday Times*, a principios de 2013, muestra con claridad el estado de la cuestión. Habida cuenta de que este estudio es en esencia similar al que pretendemos llevar a cabo aquí, su revisión nos permitirá reflexionar y obtener valiosas conclusiones que, a su vez, nos servirán de orientación a la hora de abordar nuestra tarea.

2.3.1. *The Book of Mormon*

En 1830, en Palmyra, una pequeña localidad del condado de Wayne, en el Estado de Nueva York, U.S.A., salió a la venta por primera vez un libro que iba a transformar las vidas de millones de personas en los años venideros: *The Book of Mormon: An Account Written by the Hand of Mormon upon Plates Taken from the Plates of Nephi*. En él se recoge la doctrina a partir de la cual, algún tiempo después, se constituiría el *Latter Day Saints Movement*, conocido en los países de habla hispana como *La Iglesia de Jesucristo de los Santos de los Últimos Días* y que tiene actualmente su sede en la ciudad de Salt Lake, Utah (Midgley, 1997).

De acuerdo a las creencias de los miembros de esta filiación religiosa, el contenido del libro habría sido transcrito por Joseph Smith a partir de las *Placas de Nephi*, planchas de oro grabadas con el texto original y cuya situación exacta le fue revelada por el ángel Moroni. Por su parte Smith se sirvió de varios escribas para componer el libro tal y como fue publicado finalmente. Está escrito en un estilo próximo al *Early Modern English* y desde entonces ha sido traducido a más de cien idiomas.

Como era de esperar, el *Libro del Mormón* ha generado una amplia controversia casi desde el mismo momento de su aparición. No solo su contenido, doctrina y supuestas revelaciones han sido discutidas, sino que su autenticidad también ha sido puesta en duda a partir de ciertas anomalías lingüísticas y literarias. Por su parte, los partidarios y defensores de la ortodoxia mormona han cuestionado los estudios críticos y han llevado a cabo sus propios estudios en los que recurren a las mismas herramientas que sus detractores – literarias y lingüísticas – para atestiguar su veracidad. Por desgracia, el elemento que contribuiría a despejar la mayor parte de las incógnitas que se suscitan, las Placas de Nephi originales, ha desaparecido.

De acuerdo a Joseph Smith, el ángel Moroni le pidió que se las devolviese una vez terminada la transcripción (Hardy y Parsons, 2003, Linn, 2011).

A pesar de que en otras ocasiones las circunstancias que motivan el análisis de un texto se revelan más dramáticas – notas de petición de rescate en secuestros, de despedida en casos de suicidio, o una condena a muerte basada en la confesión del inculcado, como el caso de Timothy Evans, que mencionábamos anteriormente – la cuestión de la autenticidad del *Libro del Mormón*, un texto sagrado para los fieles de esta confesión, resulta crucial para el bienestar espiritual y el modo de vida adoptado por millones de personas. Así, no es de extrañar que los análisis y los métodos empleados para llevarlos a cabo hayan sido motivo de disputas enconadas y si este ejemplo práctico de la aplicación de la Estilometría en la atribución de autoría nos resulta de interés se debe, precisamente, a que en pocas ocasiones las herramientas de análisis han sido sometidas a una revisión tan escrupulosa.

El primer estudio estilométrico del Libro del Mormon se realizó en 1980 y los autores fueron tres expertos en estadística de la Universidad Brigham Young: Wayne Larsen, Alvin Rencher y Tim Layton ([1980] 1997). Este trabajo, que suele denominarse el Estudio Larsen, tomaba como referencia la frecuencia de aparición de palabras no-contextuales, esto es, palabras funcionales, artículos, pronombres, conjunciones y preposiciones. Los investigadores establecieron bloques de texto de 2000 palabras de los supuestos autores del libro analizado que se mencionan en el propio texto sagrado, los profetas Nephi, Alma, Mormon y Moroni y sobre los que se efectuó la comparación intentando encontrar similitudes o diferencias que soportasen una u otra propuesta. Al mismo tiempo, se tomaron bloques de texto de otros autores del siglo diecinueve ajenos al texto analizado, entre ellos de Joseph Smith, como referencia comparativa del análisis discriminativo.

Los resultados obtenidos pueden resumirse en dos conclusiones: el *Libro del Mormón* fue escrito por varios autores y estos autores tienen un estilo que difiere notablemente de los autores del siglo diecinueve tomados como referencia, incluido el propio Joseph Smith. Los bloques de texto analizados muestran que existieron cuatro autores diferentes, cada uno de ellos con un estilo claramente diferenciado, ninguno de los cuales fue Joseph Smith. Por supuesto, no es posible determinar si esos cuatro escritores se corresponden con los mencionados profetas, pero al menos es posible descartar que fuesen escritos por una misma persona.

Algún tiempo después, John Hilton, físico del Laboratorio Nacional Lawrence Livermore, en California, y vinculado también a la Universidad Brigham Young, decidió reproducir los análisis estilométricos llevados a cabo en el estudio Larsen y verificar si sus resultados se correspondían con los de dicho análisis ([1982] 1997). Sin embargo, Hilton prefirió variar los parámetros de búsqueda y eligió sesenta y cinco patrones léxicos no contextuales. Este tipo de patrones – el artículo “*the*” como la primera palabra de una frase, “*and*” seguido de un adjetivo, “*and*” seguido de “*the*”, etc. – apenas se ven afectados por el contexto o el tema tratado y, de acuerdo a Hilton, revelan las preferencias inconscientes de uso léxico del escritor.

Hilton diseñó, además, un método discriminativo basado en la acumulación de “rechazos”, es decir, cuando en la comparación de dos textos, estos arrojan resultados en los que los patrones mencionados anteriormente difieren cuantitativamente, se produce un “rechazo”. Si se dan más de siete “rechazos”, se puede afirmar que los textos han sido escritos por autores diferentes. Previamente, Hilton había afinado este sistema en textos de autoría indisputada de diversos autores, todos ellos del siglo diecinueve. Además, con el fin de incrementar la fiabilidad de las mediciones, tomó fragmentos de texto de una extensión de 5000 palabras, siguiendo el razonamiento de que los

hábitos de escritura se manifiestan con más nitidez cuanto mayor sea el tamaño de los textos elegidos.

Las muestras de textos se tomaron, por un lado, de los libros atribuidos a los profetas Alma y Nephi y, por otro, de textos cuya autoría se reconoce como verificada de Joseph Smith, de Oliver Cowdery – uno de los cofundadores del movimiento religioso, junto a Smith – y de Solomon Spaulding³⁶, autor de cierta obra de ficción que supuestamente habría sido plagiada por otro de los escribas en la confección del *Libro del Mormón*.

Los resultados arrojaron similitudes significativas intra-autor y diferencias notables inter-autor. Es decir, que los textos supuestamente escritos por Alma y Nephi y los textos cuya autoría se daba por cierta de Smith, Cowdery y Spaulding, mostraban coherencia entre ellos, por lo que debían pertenecer a un único escritor, pero al compararlos unos con otros, revelaban diferencias suficientes como para afirmar que correspondían a autores distintos. Dado que, de acuerdo a dichos resultados, tanto la autoría única del *Libro del Mormón*, como la participación de Smith, Cowdery o Spaulding, era estadísticamente indefendible, Hilton concluyó que el texto analizado permanecía coherente con sus propias atribuciones internas.

El estudio de Hilton es especialmente interesante, no solo por el sistema de medición empleado, sino también porque se aseguró de “blindar” dicho sistema ante posibles objeciones. Hilton lo puso a prueba y lo calibró previamente con textos indubitados de autores de la época, como Mark Twain, y se aseguró de que el proceso de traducción no interfería con los resultados, verificando que una obra de un mismo autor, volcada al inglés por dos traductores diferentes, seguía ofreciendo una misma “huella lingüística”.

³⁶ De acuerdo a la Teoría Spaulding-Rigdon, el *Libro del Mormón* sería un plagio de un manuscrito nunca publicado y escrito por un tal Solomon Spaulding. Este manuscrito habría sido adquirido por Sidney Rigdon, quien, en colaboración con Joseph Smith, se habría servido de él para escribir el *Libro del Mormón* (Linn, 2011).

Algún tiempo después, en 1985, David Holmes, del Bristol Polytechnic, realizó su propio análisis estilométrico, en esta ocasión tomando como valor de referencia la diversidad léxica, estimando que “es una medida particularmente efectiva para discriminar entre autores” (1991). Holmes midió las palabras que aparecían una única vez, las que aparecían dos veces y una combinación de parámetros (Sichel, 1986) para analizar la distribución de las repeticiones de palabras. Los resultados obtenidos indicaban que los textos atribuidos a los profetas Nephi, Mormon, Lehi, Jacob y Moroni y los de Joseph Smith no eran estilométricamente diferentes, lo que contradecía los resultados obtenidos por Larsen y Hilton, y concluía que no había evidencia de múltiple autoría.

Sus resultados fueron contestados por Bruce Schaalje, Matthew Roper y Paul J. Fields (2012), sobre la base de que Holmes había utilizado un razonamiento viciado: Que él no hubiese encontrado diferencias en cuanto a la riqueza léxica entre unos textos y otros, no significaba que no las hubiese de otro tipo; significaba tan solo que él no las había encontrado. Holmes habría realizado una sobre-generalización. Por otro lado, Schaalje, Hilton y Archer (1997) mostraron posteriormente que la riqueza léxica es un pobre elemento discriminador de autoría. Al aplicar este método en textos indubitados de Samuel Clemens (Mark Twain) y Samuel Johnson, los índices de acierto alcanzaban solo el 23%, mientras que empleando los sistemas anteriores – las palabras funcionales aisladas y en combinación – esta cifra subía hasta 96% y 92%.

Tal vez el estudio más reciente sea el “estudio Jockers”, en el que colaboraron Matthew Jockers, Daniela Witten y Craig Criddle (2008). Se trata de un estudio que aplica estadísticas bayesianas³⁷, en concreto el método

³⁷ La inferencia bayesiana (Thomas Bayes, 1702-1761) es una técnica de análisis estadístico basada en la existencia de una hipótesis previa, a la cual se le asigna una tasa de probabilidad que irá modificándose a medida que se dispongan de nuevos datos. Estos datos, que pueden recolectarse por muy diferentes medios, irán alterando la fiabilidad de la hipótesis inicial, a medida que la evidencia en un sentido o en

denominado “Nearest Shrunk Centroid Classification” (NSC), utilizado con frecuencia en la investigación médica y psicológica. En esencia este método busca reducir la incertidumbre de una predicción acumulando información de diversas fuentes. El resultado aparece como la acumulación de una serie de valores próximos entre sí (“cluster”), que en este caso definen el estilo de un escritor. Cuanto más próximo aparezca el *cluster* de un texto dubitado del *cluster* de un texto indubitado más probabilidades habrá de que pertenezcan al mismo escritor. En el “estudio Jockers” sobre el *Libro del Mormón* los investigadores establecieron un panel cerrado de posibles candidatos: Solomon Spaulding, Sidney Rigdon, Oliver Cowdery y Parley P. Pratt.

El rasgo estilístico elegido en esta ocasión fueron las 110 palabras más frecuentes de entre las que se encuentran en el texto examinado y que aparecen también, como mínimo una vez, en los textos de los autores candidatos.

Sus conclusiones son que el texto analizado fue escrito por varios autores del siglo diecinueve y señalan como más probables a Rigdon y Cowdery – “quienes tenían los motivos, los medios y la ocasión” (Jockers, Witten y Criddle, 2008: 487) – a partir de una obra original de Spaulding. Como argumento para dejar fuera del panel de candidatos a Joseph Smith, alegan que dado que este empleó con frecuencia secretarios o escribas, los textos disponibles no son fiables.

Este último punto es una de las objeciones que Schaalje, Roper y Fields ponen a sus resultados (2012). Ya que Smith fue el traductor, él es el

otro se acumula (Joyce, 2008). Por consiguiente, conlleva un cierto grado de subjetividad, dado que las primeras estimaciones no están basadas más que en “grados de creencia” o “intuiciones”, lo que no tiene por qué ser necesariamente menos fiable: si, por ejemplo, establezco la hipótesis de que si salto por ventana de mi habitación, caeré hacia el suelo, sin duda puedo asignarle un grado de probabilidad muy elevado de que se cumpla, aun antes de haber comenzado a realizar las pruebas para verificarla. No todos los expertos en estadística están de acuerdo con este sistema, sin embargo, pues algunos afirman que las creencias iniciales pueden influir en la recolección de las evidencias, como bien podría ser el caso en un estudio que pretendiese demostrar, por ejemplo, la existencia o inexistencia de Dios.

candidato más probable, arguyen y en cualquier caso, un panel cerrado de posibles candidatos solo debe plantearse cuando existen evidencias de que el autor se encuentra entre ellos, algo que no sucede en esta ocasión.

Esta argumentación no carece de peso; si proponemos un número concreto de candidatos y establecemos como el autor más probable aquel cuyos textos indubitados se sitúen más cerca del estilo del texto sometido a examen, podemos estar proporcionando un “falso positivo”, ya que nada nos garantiza que fuera de este grupo cerrado de candidatos encontremos a otro escritor que se aproxime todavía más al estilo del texto de referencia. En la práctica, sin embargo, factores contextuales e históricos eliminan una buena parte de los posibles candidatos y esto es lo que hacen los autores del Estudio Jockers, lo que les lleva a afirmar, como hemos dicho más arriba, que quienes ellos señalan tenían “los motivos, los medios y la ocasión”.

Por otro lado, cuando en un panel de candidatos cerrado elegimos como “más probable” al que esté “más cerca” del texto de referencia, esto sigue siendo un valor relativo, porque la distancia entre ambos, en términos reales, puede ser, a pesar de todo, muy grande, tal y como afirman Schaalje, Roper y Fields. Para que el método NSC sea de utilidad en estilometría, consideran que el investigador debe saber con un grado de certeza razonable que no existe ningún otro posible autor, fuera de los incluidos en el panel de candidatos. Un ejemplo de este último escenario lo constituye el caso *The Federalist Papers*.

2.3.2. *The Federalist Papers*

En los días previos a la ratificación de la Constitución de los Estados Unidos, en 1787, tanto defensores de la misma como opositores hicieron públicas sus opiniones divergentes. De entre los primeros, tres nombres

destacan por su importante aportación al debate: James Madison, Alexander Hamilton y John Jay³⁸. Estos tres políticos y estadistas escribieron un total de ochenta y cinco ensayos promoviendo el apoyo a la Constitución; esos ensayos de un elevado valor histórico, literario y lingüístico, han llegado a conocerse como *The Federalist Papers*.

Publicados en un primer momento en periódicos locales, fueron más tarde recopilados e incluidos en dos volúmenes que vieron la luz en 1788, aunque el nombre por el que son conocidos actualmente en conjunto no les sería atribuido hasta el siglo XX. La autoría de los textos se mantuvo en un primer momento en el anonimato, puesto que los tres escritores emplearon un seudónimo común, “Publius”, si bien su posible identidad no tardó en circular entre los lectores en forma de rumor. En la actualidad se conoce con certeza razonable la autoría de setenta y tres ensayos³⁹ y durante algún tiempo se mantuvo la incertidumbre al respecto de doce de ellos. Douglas Adair, en “The Disputed Authorship of the Federalist Papers”, asegura:

“When Edward Mead Earl in 1937 brought out the most recent edition of *The Federalist* he marked twelve essays as still awaiting positive proof of authorship. The disagreement has been since the beginning cause célèbre; a mystery quite as intriguing as the authorship of *Junius*⁴⁰; and one far more worthy of scholarly attention.” (Adair, 1944: 97).

³⁸ James Mádison, Jr. (1751 – 1804) fue un destacado estadista, que sería reconocido por la historia como “el padre de la Constitución”. Madison llegaría a desempeñar el cargo de cuarto presidente de los Estados Unidos. Alexander Hamilton (1755 – 1804) promotor de la Constitución desde su puesto como asistente de George Washington, fue también el fundador del primer partido político. John Jay (1741 – 1829), uno de los Padres Fundadores, fue igualmente el primer Jefe de Justicia de los Estados Unidos (Bradford y Kirk, 1994).

³⁹ Al parecer la primera fuente de información al respecto fue el propio Alexander Hamilton, quien facilitó a su abogado una relación de la autoría de cada ensayo en los días previos al fatal desenlace del duelo con pistola entre él y Aaron Burr, que se resolvió con el fallecimiento de Hamilton. Algún tiempo después, Madison proporcionó su propia lista de autoría, en la que se atribuía a sí mismo veintinueve ensayos y explicaba la diferencia entre ambas listas “debido a la premura y las especiales circunstancias” bajo las que fue escrita la lista de Hamilton (Adair, 1944: 122).

⁴⁰ Entre 1769 y 1772, el periódico londinense *Public Advertiser* recibió y publicó una serie de misivas de contenido político y social, firmadas con el seudónimo de “Junius”. Por el estilo en que están escritas, se presume que el autor podría pertenecer a las clases más altas y educadas de la sociedad y desde entonces

En 1964 Frederick Mosteller y David L. Wallace llevaron a cabo el que está considerado el primer estudio estilométrico en el que se aplicaban las modernas técnicas de análisis estadístico. Según sus propias palabras, su intención fue presentar un estudio sistemático y comparativo entre métodos bayesianos y clásicos aplicados a problemas discriminativos. Para Mosteller y Wallace los métodos tradicionales para determinar diferencias de estilo tomando como referencia la longitud de las frases no resultaba adecuada, debido a la gran similitud que en este aspecto ofrecían los escritos de Madison y Hamilton⁴¹.

Los investigadores recurrieron por tanto al cómputo de palabras y para ello midieron los índices de aparición de determinados ítems en bloques de una extensión de 200 palabras. Mosteller y Wallace buscaron en un primer momento las preferencias de cada autor por determinados sinónimos; sin embargo los pares de sinónimos escaseaban por lo que buscaron otro tipo de palabras que apareciesen con una frecuencia relativa, pero en proporciones diferentes y se centraron en preposiciones, conjunciones y artículos (Juola, 2008). Utilizando el teorema de Bayes, emplearon escalas numéricas de probabilidad para “expresar la certeza de hipótesis iniciales acerca de la autoría de los distintos candidatos” (Holmes, 1985: 333).

Para determinar qué palabras habrían de servir como elementos de referencia, hicieron numerosas pruebas y ensayos con textos cuya autoría se conocía con certeza hasta establecer la lista de las 30 palabras funcionales que tomarían como marcadores para llevar a cabo su análisis. Para el estudio principal, las palabras seleccionadas se agruparon en cinco clases, de entre las

hasta el día de hoy han sido sometidas a análisis con el fin de intentar determinar quién se ocultaba bajo este seudónimo. Tanto por su elegante estilo como por su denuncia de la corrupción entre los gobernantes, “Junius” adquirió durante algún tiempo una gran popularidad entre los lectores del *Advertiser*. La opinión de los estudiosos se decanta hacia el político Sir Philip Francis como el más probable autor de las cartas (Katritzky, 1996).

⁴¹ La longitud media de las frases de James Madison y de Alexander Hamilton es de 34,55 y 34,59 palabras, respectivamente, lo que es insuficiente para llevar a cabo una discriminación de autoría (Smith, 2006).

cuales el de las palabras con un alto índice de aparición resultó ser el más útil para sus propósitos. El trabajo de Mosteller y Wallace fue el primero en recurrir a las palabras funcionales como uno de los parámetros más fiables a la hora de establecer la autoría (Koppel, Schler y Argamon, 2009).

Este sistema se apoyaba en una clasificación bayesiana de los documentos y suele conocerse por el nombre de *Análisis de Aproximación Multivariable* (“*Multivariate Analysis Approach*”). Examinando las frecuencias de un conjunto de términos no-contextuales y asignando a los resultados determinadas posiciones en el espacio (gráficas), se atribuye la autoría de un texto a aquel escritor cuya posición en la gráfica se encuentre más cerca del texto dubitado. En esencia viene a ser el sistema aplicado por el “Estudio Jockers” en el análisis del *Libro de Mormón*, solo que en este caso, y a diferencia del anterior, el panel de candidatos incluía sin lugar a dudas al autor de los textos dubitados. En realidad, los documentos reunían las condiciones ideales para un estudio de este tipo, se trataba de un ejemplo perfecto de la situación de “panel cerrado” de candidatos, con suficiente material indubitado como para realizar todas las pruebas necesarias y que versaba, además, sobre el mismo tema, había sido escrito en el mismo momento histórico y había sido publicado en el mismo medio.

En su estudio, los investigadores emplearon igualmente otros sistemas de selección de palabras, como por ejemplo, la elaboración de una discriminación lineal a partir del “peso” de cada palabra, en relación con su frecuencia de aparición en la mitad de cada uno de los textos no cuestionados. El resultado se calibró aplicándolo sobre una de las dos mitades de los textos indubitados y posteriormente se verificaba sobre la otra mitad, antes de ponerlos en práctica con los textos sobre cuya autoría se dudaba.

En las conclusiones de su estudio Mosteller y Wallace señalaban a James Madison como el autor más probable de los doce ensayos disputados,

un resultado que no ha sido discutido desde entonces. Este trabajo se ha convertido con los años en una referencia y su metodología, en concreto el empleo de las palabras funcionales como el rasgo que con más probabilidad contribuye a señalar la autoría, ha sido puesta en práctica en numerosas ocasiones posteriores. Estudios psicológicos independientes de la atribución de autoría han corroborado su utilidad.

“The words we use in natural language reflect our thoughts and feelings in often unpredictable ways. They also reveal a tremendous amount of information about our social interactions and personality. Function words, in particular, carry an array of psychological meanings and set the tone for social interactions.” (Chung y Pennebaker, [2003] 2007: 355).

Las dos razones principales son, en primer lugar, que al carecer de carga semántica no se hallan asociadas a contextos concretos y su distribución es muy similar sea cual sea el tema tratado; en segundo lugar, que resulta muy improbable que su utilización, tanto en el habla como en la escritura, sea controlada conscientemente y son “con toda probabilidad (..) un mero reflejo de la actividad cognitiva subyacente” (Chung y Pennebaker, [2003] 2007: 356), lo que también contribuye a desenmascarar engaños intencionados.

2.3.3. *The Cuckoo's Calling*

En abril de 2013, la editorial inglesa Little, Brown Book Group publicó la novela *The Cuckoo's Calling* firmada por Robert Galbraith, un escritor debutante y al que los editores describían como “investigador retirado del cuerpo Royal Military Police, que desde el 2003 trabaja en el campo de la seguridad civil”. La novela describe las indagaciones que Cormoran Strike,

detective privado y ex militar, lleva a cabo sobre el aparente suicidio de una supermodelo, por encargo de un hermano de la fallecida.

La crítica especializada la recibió como una excelente opera prima, ensalzando tanto la trama como la construcción de los personajes (Burke, 2013; Finch, 2013; *Publishers Weekly*, 1993). En *The Telegraph*, Jake Kerridge (2013) le otorgaba cuatro estrellas, de cinco. En cambio, las ventas fueron modestas: 1.500 ejemplares en edición impresa y otros siete mil en audio libros y *e-book*. Sin embargo, en julio de ese mismo año, la prensa publicaba la noticia de que Robert Galbraith era en realidad Joanne Rowling (Bury, 2013), la conocida autora de los libros sobre el mago Harry Potter, que acostumbra a firmar como J. K. Rowling⁴². El efecto que la noticia tuvo en el mercado editorial fue notable: durante las semanas que siguieron las ventas de la novela se elevaron hasta un 150.000 % (Spence, 2013).

El diario *Sunday Times* había recibido, algún tiempo atrás, a través de Twitter la confidencia de que la novela podría haber sido escrita en realidad por Rowling. Bien conocida por el público en general, Rowling se ha convertido en una de las escritoras de más éxito de la última década. Sus novelas juveniles, llevadas al cine en una saga de siete capítulos, han acompañado a toda una generación de adolescentes.

Lo cierto es que existían ya antes algunos indicios de que quien firmaba *The Cuckoo's Calling*, Robert Galbraith, podría ser un seudónimo. Galbraith, por ejemplo, describía las vestimentas femeninas de forma

⁴² Por otro lado, el incidente resulta interesante porque revela las complejas reglas que rigen las relaciones entre escritores, críticos y público lector. Apenas el año anterior, Rowling había publicado otra novela, *The Casual Vacancy*, en este caso bajo su auténtico nombre. Esta obra, que resultó un éxito en lo que a ventas se refiere, fue, en cambio, maltratada por la crítica literaria (Oppenheimer, 2013). Para algunos, el conocido nombre de la escritora, asociado a sus éxitos comerciales anteriores, condicionó tanto el aprecio de los lectores, como el criterio con que la novela fue valorada por los críticos. Justo lo contrario de lo que sucedió con la aparición de *The Cuckoo's Calling*: mientras el autor pareció ser un desconocido, la crítica la ensalzó, pero las ventas fueron casi anecdóticas. En cambio, al salir a la luz el nombre de su auténtica autora, los lectores, probablemente muchos de ellos fans de Rowling, compraron la novela disparando las ventas. Los críticos, en cambio, difícilmente podían para entonces dar marcha atrás en sus calificaciones previas.

especialmente precisa, lo que para algunos sugería que era una mujer la escritora (Oppenheimer, 2013). Por otro lado, Rowling había manifestado anteriormente su intención de escribir una historia de detectives y se conocía de antemano que el editor de su siguiente libro sería David Shelley, quien acostumbra a trabajar con escritores especializados en el género negro (Oppenheimer, 2013). A partir de esta información, el *Sunday Times* decidió encargar a Patrick Juola, de la Universidad Duquesne, en Pittsburgh, un estudio sobre la novela y un dictamen al respecto de si podría o no haber sido escrita por Rowling.

Juola (2013a, 2013b) dispuso de cuatro textos indubitados contra los que someter a contraste la novela cuestionada: *The Casual Vacancy*, escrito por la propia Rowling, *The Private Patient*, de P.D. James, *The Zita Society*, de Ruth Rendell y *The Wire in the Blood*, de Val McDermid. Los textos le fueron proporcionados en formato electrónico y la herramienta informática que Juola utilizó fue el programa JGAAP, diseñado por él mismo.

Para el análisis, Juola dividió *The Cuckoo's Calling* en once fragmentos de mil líneas cada uno, lo que equivale a unas diez mil palabras, aproximadamente, y tomó en consideración cuatro variables o marcadores. El primero fue la longitud de las palabras – medida en número de letras contenidas en cada palabra – y los porcentajes con que aparece cada longitud considerada. A partir de ahí, estableció una medida de similitud con 0.0 para la identidad absoluta e incrementos numéricos cuanto mayor fuese la diferencia.

El resultado fue que de los once fragmentos, seis se aproximaban al libro indubitado de Rowling – *The Casual Vacancy* – y cinco al de James. En cuanto a los otros dos, quedaban demasiado lejos como para considerarlos.

El siguiente marcador empleado fue la relación de las cien palabras más frecuentes, de nuevo clasificadas por porcentajes de aparición, incluyendo tanto palabras de contenido semántico, como funcionales. En esta ocasión cuatro fragmentos del texto cuestionado se asemejaban al indubitado de Rowling, cuatro al de McDermid y otros tres se acercaban por igual al de James y al de Rendell.

Para los dos marcadores restantes, Juola recurrió a dos n-gramas diferentes, el primero de los cuales lo formaban 4-gramas, compuestos de cuatro caracteres adyacentes, incluyendo los espacios entre palabras y tomando letras pertenecientes a dos palabras, si así coincidía en el cómputo. En este caso era McDermid quien obtenía mayor similitud, con ocho fragmentos coincidentes, mientras que Rowling obtenía solo tres.

El segundo n-grama estaba formado por dos palabras consecutivas y en este caso Rowling obtenía nueve fragmentos que se le aproximaban, McDermid obtenía dos y ninguna otra autora se acercaba.

A partir de estos resultados inconclusos todos ellos tomados por separado, Juola observa que, en conjunto, apuntan de forma convincente hacia Rowling. Ningún resultado señala a Rendell como posible autora, mientras que solo uno sugiere a James. Juola admite que todo ello no *demuestra* que Rowling sea la autora del libro cuestionado, sino que de las cuatro autoras candidatas, la más *probable* es J.K. Rowling. ¿Podría ser que alguien, otra persona, imitando el estilo de Rowling con inusual destreza, hubiese escrito *The Cuckoo's Calling*? Sí, es posible, admite Juola. Sin embargo, aquí intervienen los factores extralingüísticos para señalar a la escritora “sospechosa” como la más probable, que sumados a los indicios obtenidos por medio del análisis estilométrico, resultaron lo suficientemente convincentes como para que los editores de *Sunday Times* se dirigiesen al agente de Rowling y le preguntasen si había sido o no su representada quien había

escrito el libro. Al día siguiente el representante de Rowling hacía público un comunicado en el que la escritora reconocía ser la autora de la novela⁴³.

El razonamiento que Juola aplica a la hora de emitir su juicio resulta de gran interés, pues revela con claridad las posibilidades reales que ofrece el método con que lleva a cabo su análisis:

“We can do some crude statistics about the likelihood that a randomly chosen author would have a style that similar to Rowling, and by extension how strong this suggestion really is. Out of four candidates, Rowling was consistently #1 or #2 (i.e., in the more similar half) of each feature chosen; it's therefore only 50/50 that a randomly chosen author would be in the more similar half. With four studies (handwaving independence assumptions), there's therefore only one chance in 16 that the person would "pass" all studies as being similar to Rowling. If we needed a stronger suggestion, we could easily gather more data, more distractor authors, or simply run more experiments on different variables.” (Juola, 2013a).

Es cierto que de tener que elegir entre las dos escritoras que obtienen mejores resultados, la posibilidad de que Rowling sea la autora es del 50 % y que, dado que son cuatro las candidatas y cuatro también las pruebas realizadas, cada una de forma independiente de las otras tres, la probabilidad de autoría a partir de los resultados en conjunto aumenta hasta la proporción de 16: 1. Sin embargo, para que esta argumentación se sostenga, es preciso

⁴³ Rastrear el origen de la filtración que puso en movimiento a los elementos de la trama hubo de requerir una concienzuda investigación en sí misma. Al parecer, la fuente fue Chris Gossage, miembro del bufete de abogados Russells, que representan a J.K. Rowling desde hace años, quien reveló a la mejor amiga de su mujer, Judith Callegarim, que Galbraith era en realidad un seudónimo de Rowling. Un *tweet* colgado de la red social por Callegarim situó a *Sunday Times* sobre la pista que llevó finalmente a desenmascarar a la autora de la novela (Gladdis, 2013). Rowling inició una demanda judicial contra el bufete Russells, pero desistió tras llegar a un acuerdo, según el cual, la empresa haría una donación a la ONG *Soldier's Charity*, que se ocupa de prestar asistencia a militares retirados y a sus familias (Waller-Davies, 2013). Por otro lado, la escritora se ha comprometido a hacer otro tanto con los beneficios que produzca la novela desde el momento en que fue revelado su nombre y por un espacio de tres años (Stansall, 2013), sin duda para disipar las sospechas de que todo había sido un montaje destinado a incrementar las ventas.

que el panel de candidatos sea cerrado y el analista tenga la certeza de que la autora se encuentra entre las cuatro seleccionadas. En el presente estudio, tres de las candidatas fueron elegidas a partir de la condición de que fuesen mujeres y autoras de novelas del género negro. La cuarta candidata, la “sospechosa”, se incluyó a partir de una serie de indicios extralingüísticos – a saber: que Rowling había hablado de su intención de escribir una novela de detectives, que su nuevo agente literario se especializaba en trabajos de este género y un cotilleo publicado en Twitter – ninguno de ellos demasiado consistente.

El “final feliz” de la investigación lo constituye la admisión de autoría de Rowling, pero en el caso de que no hubiese sido así – como podría ocurrir en el supuesto de que el autor “sospechoso” hubiese fallecido tiempo atrás – las disputas al respecto de los resultados se hubiesen extendido en las publicaciones especializadas a lo largo de años, tal y como el propio Juola reconoce.

De haber trabajado con solo dos candidatos, escogidos al azar y que hubiesen obtenido resultados similares, las probabilidades de acierto, no serían desde luego del 50%, sino que dependerían del total de la población a la que considerásemos potencialmente autora del texto en cuestión, de la misma manera que, si compramos un billete de lotería, las opciones son “que nos toque” o “que no nos toque”, pero la probabilidad de que “nos toque” depende de la cantidad de números que toman parte en el sorteo.

Por otro lado, los resultados muestran que ninguno de los sistemas aplicados es concluyente, como lo sería aquel en el que la totalidad de los once fragmentos señalaran a Rowling, y solo a Rowling, como la autora del libro cuestionado. Juola propone, para obtener una mayor *convicción*, aumentar los factores que intervienen en el análisis – trabajando con mayor número de autores o realizando pruebas con nuevos marcadores de estilo –

multiplicando así la fiabilidad del resultado. En la práctica, tan solo nuevas pruebas con nuevos marcadores tendrían influencia sobre la probabilidad real de acercarnos al autor, dado que si añadiésemos más candidatos – a los que Juola llama apropiadamente “distracciones” – escogidos al azar entre mujeres escritoras de novelas de detectives, aumentaría el grado de certeza; pero, para saber en qué proporción, necesitaríamos conocer con precisión el número total de escritoras de novelas de detectives. Además, siempre nos quedarían otras posibilidades, en absoluto descabelladas, como que el autor fuese un hombre o una mujer y con experiencia previa o no en dicho género.

De hecho la necesidad de introducir otros autores candidatos proviene sobre todo de la incapacidad de los sistemas de análisis para producir un perfil estilístico consistente y reconocible del autor “sospechoso”, pues de ser así, solo tendríamos que contrastar el perfil del candidato con el del autor del libro dubitado – sea este quien sea – y ver si coinciden o no. Sin embargo, esta importante cuestión la trataremos de nuevo, más adelante y de forma más extensa, cuando describamos el sistema que aplicaremos en el caso que nos ocupa en este trabajo.

3. Carlos Castaneda:
Presentación del autor y de su obra

3.1. El relato de un aprendizaje

En junio de 1968, California University Press publica *The Teachings of Don Juan: A Yaqui Way of Knowledge*, un volumen firmado por Carlos Castaneda, joven estudiante de antropología en esa misma universidad. El libro se presenta como la transcripción del cuaderno de campo en el que Castaneda ha ido recogiendo los resultados de su investigación sobre ciertas prácticas propias de la cultura de los indios mexicanos.

De acuerdo a lo narrado en el libro, a principios de 1960, Castaneda habría llegado a la región de Sonora, situada en el noroeste de México, buscando informantes que le proporcionasen datos sobre el uso de plantas medicinales entre los indios de la región. Sin embargo, un encuentro casual con quien más tarde resultaría ser un brujo yaqui terminaría convirtiéndole en aprendiz de un antiguo sistema de conocimientos.

“At first I saw don Juan simply as a rather peculiar man who knew a great deal about peyote and who spoke Spanish remarkably well. But the people with whom he lived believed that he had some sort of "secret knowledge", that he was a "brujo". The Spanish word brujo means, in English, medicine man, curer, witch, sorcerer. It connotes essentially a person who has extraordinary, and usually evil, powers.” (1968: 3).

Este brujo, a quien Castaneda llama “don Juan”, inicia al joven estudiante en una serie de prácticas que le permitirán incrementar su “poder personal” y acceder a una realidad alternativa, de la que todos somos ignorantes, pero que convive con esta que percibimos en nuestra vida

cotidiana. El proceso de enseñanza, descrito en este primer libro, tiene lugar durante las frecuentes visitas que desde entonces y hasta 1965 Castaneda realiza a México para encontrarse con su maestro. El autor, a la vez que sigue las indicaciones que don Juan le hace cada vez que están juntos, intenta conservar su perspectiva como antropólogo y toma constantemente notas de todo lo que sucede, incluidos los diálogos con su mentor.

“I believe this method of conducting the apprenticeship prevented the training from being successful, because it retarded the advent of the full commitment I needed to become a sorcerer. Yet the method was beneficial from my personal standpoint in that it allowed me a modicum of detachment, and that in turn fostered a sense of critical examination which would have been impossible to attain had I participated continuously, without interruption. In September 1965, I voluntarily discontinued the apprenticeship.” (1968: 5).

A pesar del calificativo de “brujo” que Castaneda le atribuye en un primer momento, don Juan se considera a sí mismo “un hombre de conocimiento”, heredero de un antiquísimo linaje tolteca de hombres que han obtenido la sabiduría suficiente como para trascender la realidad de este mundo. El proceso de iniciación, sin embargo, resulta tan aterrador para el estudiante de antropología, que termina por alejarse de su maestro y de las prácticas de este.

Una vez abandonado su aprendizaje, Castaneda intenta dar forma a sus notas de campo y clasificar de forma sistemática la información que ha ido recogiendo durante ese tiempo y que él clasifica en: uso de plantas alucinógenas, procedimientos y fórmulas empleadas en brujería, adquisición y manipulación de objetos de poder, uso de plantas medicinales y canciones y leyendas.

“Reflecting upon the phenomena I had experienced, I realized that my attempt at classification had produced nothing more than an inventory of categories; any attempt to refine my scheme would therefore yield only a more complex inventory. That was not what I wanted. During the months following my withdrawal from the apprenticeship, I needed to understand what I had experienced, and what I had experienced was the teaching of a coherent system of beliefs by means of a pragmatic and experimental method. It had been evident to me from the very first session in which I had participated that don Juan's teachings possessed an internal cohesion.” (1968: 5).

Carlos Castaneda organiza su libro en dos secciones; la primera de ellas consiste en una selección de sus anotaciones en las que describe sus andanzas – su consumo de plantas alucinógenas, su acceso a estados de realidad no-ordinaria y sus conversaciones con su maestro don Juan – y en la que proporciona su versión subjetiva de los acontecimientos percibidos. En la segunda parte, que lleva el título de *A Structural Analysis*, elabora un ensayo escrito en un estilo que podríamos llamar “académico” y en el que intenta dar forma a todo lo anterior de acuerdo a las teorías antropológicas. En conjunto, el texto fue presentado como la tesis del máster en antropología que Castaneda cursaba en ese momento.

Tan pronto como sale a la luz, el libro se convierte en un éxito de ventas. Las andanzas del joven Castaneda complacen a la vez a la crítica especializada y a los lectores, ávidos de narraciones que confirmen la existencia de una realidad diferente, invisible, pero cercana y a la que es posible acceder mediante el uso de ciertas sustancias, que, en esos momentos, se están poniendo muy de moda. Los años sesenta del pasado siglo fueron una época de profundos cambios en los terrenos social y cultural, en los Estados Unidos. Tom Wolfe, en su antología de ensayos y reportajes publicados en esos años, *The Pump House Gang* (1968: 3), asegura que fue “la inesperada bonanza económica surgida tras la Segunda Guerra Mundial la que propició

una nueva actitud hacia la vida” entre sus compatriotas. La inyección de dinero en las clases media y baja permitió a la gente expresar su creatividad como nunca antes lo había hecho, provocando por todo el país lo que él llamó una “explosión de felicidad”.

Surgieron estilos de vida alternativos al *American Way*, especialmente en California, y de todos ellos tal vez el más llamativo fue el movimiento *hippie*, que mezclaba el misticismo de las filosofías orientales con un consumo desinhibido de sustancias psicotrópicas. Los experimentos con drogas psicodélicas de los académicos Timothy Leary y Richard Alpert, en la Universidad de Harvard y, más tarde, su defensa del consumo de LSD como medio de provocar estados de conciencia alterados conmocionaron a la opinión pública, pero despertaron un enorme interés entre las generaciones más jóvenes y entre los intelectuales de todo el país. El momento, desde una perspectiva social y cultural, era perfecto para la aparición de un texto que, como sucedía con el libro de Castaneda, diese validez a estos nuevos planteamientos vitales. La propuesta del joven Castaneda proporcionaba un soporte filosófico y étnico al consumo de drogas y al afán por alcanzar otra realidad, otros mundos, que, ajustándose a la frase que se le atribuye al poeta Paul Eluard, conviven con este.

Tres años más tarde aparece *A Separate Reality: Further Conversations with Don Juan* (1971), libro en el que Castaneda retoma su relación con don Juan y en el que describe sus esfuerzos por aclarar algunos eventos de los que había sido testigo anteriormente y por obtener una explicación que los justifique desde un punto de vista racional. Castaneda reinicia así su aprendizaje, acompaña a su maestro por los desiertos mexicanos y una vez más presencia y se ve obligado a tomar parte en hechos que desafían la lógica y que él atribuye a “trucos” del brujo yaqui o a los efectos del consumo de las plantas alucinógenas, a pesar de que buena parte de dichos sucesos tienen lugar durante períodos de vigilia consciente y sobria.

En la narración aparece por primera vez otro brujo, don Genaro, compañero de don Juan, pero más terrorífico que este, si cabe. Entre ambos intentarán enseñarle a “ver”, un acto propio de un “hombre de conocimiento” que le permite “percibir la energía que fluye a través del universo”.

Castaneda mantiene en todo momento una actitud escéptica, cuestionando cada evento extraordinario, lo que da lugar a diálogos con don Juan en los que este intentará hacer entender a su aprendiz la estructura de una realidad diferente a la que está acostumbrado y a la que se aferra desesperadamente. A pesar de todos sus esfuerzos, Castaneda no encontrará más explicaciones a lo que presencia que las que su maestro le proporciona, por muy insatisfactorias que le resulten. Tras diez años de aprendizaje, sigue negándose a considerar “reales” los estímulos que percibe, a pesar de que ya no puede sostener sus viejos criterios acerca de qué es real y qué no, y afirma que bajo la presión de los actos de don Juan y don Genaro “su mente ha entrado en un callejón sin salida” (1971: 301).

Este segundo volumen, narrado igualmente en primera persona y con el formato de un diario de campo, acrecienta su prestigio tanto en los medios académicos, como entre el público lector, aun cuando comienzan a surgir voces que expresan ciertas dudas acerca de su valor como documento antropológico. En efecto, Castaneda no ha aportado hasta el momento evidencia alguna de sus encuentros con don Juan, ni ha mostrado a nadie sus notas personales.

A pesar de todo ello su siguiente libro, *Journey to Ixtlan* (1972) es presentado en la Universidad de California (UCLA) como la tesis que le proporciona un doctorado en antropología. En su narración, Castaneda, que mantiene su actitud escéptica y racionalista, se ve obligado a reconsiderar la importancia que el consumo de plantas psicotrópicas tiene dentro de su proceso de aprendizaje. Dichas plantas – peyote o “mezcalito”, *Datura Innoxia*

o “yerba del diablo” y los hongos *Psilocybe* o “humito” – no son, al contrario de lo que él había creído hasta entonces, el único camino para acceder a la percepción de esa realidad alternativa de la que su maestro le habla.

“Don Juan had stated that in order to arrive at “seeing” one first had to “stop the world.” “Stopping the world” was indeed an appropriate rendition of certain states of awareness in which the reality of everyday life is altered because the flow of interpretation, which ordinarily runs uninterruptedly, has been stopped by a set of circumstances alien to that flow.” (1972: xiii).

Don Juan explica a Castaneda que existen varias etapas en el proceso de aprendizaje – cazador, guerrero y, por último, hombre de conocimiento – que deben ser superadas sucesivamente. Cada una de ellas conlleva la necesidad de incrementar su “poder” personal mediante técnicas, objetos, animales o entidades inmateriales que deben ser dominadas. La narración finaliza con don Genaro contando una etapa concreta de su aprendizaje, que da título al libro. Durante su propio proceso de adquisición de “poder” y de la capacidad para “ver” el mundo a la manera de un hombre de conocimiento, a don Genaro, un día, en el camino de regreso a casa después de una larga estancia en el desierto, todas las personas con las que se cruza se le aparecen como fantasmas, como si careciesen de sustancia real y los lugares que antaño le habían sido familiares le resultan ahora desconocidos. Durante algún tiempo vaga perdido, hasta que por fin comprende que una vez que “inicias el camino del conocimiento” ya no hay vuelta atrás, ya nada es lo mismo. Por eso don Genaro ya nunca podrá volver a su casa en Ixtlán, lo mismo que Castaneda no podrá regresar a la ciudad de Los Ángeles ni y a la vida que ha llevado hasta entonces, antes de comenzar su aprendizaje.

Tales of Power (1974), el cuarto libro, es un repaso a diversos encuentros y conversaciones mantenidos con don Juan y don Genaro, situados

cronológicamente varios años atrás. Esta recapitulación examina diversos consejos que el maestro de Castaneda le transmite y que ayudan a perfilar el carácter y el ánimo con que un “guerrero” afronta las vicisitudes que le ocurren en su vida cotidiana. Hacia el final del libro, Carlos Castaneda finaliza su aprendizaje como “hombre de conocimiento” y, al ponerse el sol, salta desde lo alto de un abismo en medio del desierto y desapareciendo en la oscuridad.

Varios conceptos resultan fundamentales en estos cuatro primeros libros. El primero de ellos es el de “borrar la historia personal”. En palabras de De Mille:

“A warrior erases his personal history. He does not tell his name, his origin, or his previous occupation. Sometimes he forgets those things himself. Having no idea what to expect of him, other people cannot pin him down with their thoughts.” (De Mille, 1976: 127).

Esta fórmula para permanecer fuera del alcance de otros enlaza directamente con el segundo de dichos conceptos: “*ser inaccesible*”. El guerrero debe mantenerse fuera del alcance de quienes le rodean y no permitir que le “*sitúen*” en un momento o lugar determinado. Lo cierto es que Castaneda se ajusta a estas premisas en gran medida, manteniendo oculta su vida privada, incluso a las personas que están más cerca de él.

El tercer concepto es el de “parar el mundo”, lo que, de acuerdo a las enseñanzas de don Juan, implica suspender el diálogo interno y dejar de emitir juicios sobre lo que nos rodea. Al hacerlo así, el adepto deja de “construir” la realidad de acuerdo a una serie de ideas preconcebidas, ideas que son transmitidas a todas las personas desde su infancia a través de sus padres, sus profesores y la sociedad en general.

Estos cuatro libros constituyen, por así decirlo, la primera de las dos partes en que se divide la obra de Carlos Castaneda. En la segunda, compuesta de otros ocho volúmenes, Castaneda se embarca en una serie de aventuras místicas, más extraordinarias todavía que las narradas hasta entonces. Tanto don Juan, como don Genaro, han desaparecido trasladados a otro plano de la existencia, por lo que Castaneda pasa a ocupar su lugar y sus siguientes libros describen los procesos de enseñanza a que somete, ahora en el papel de maestro, a sus discípulos.

Los libros que constituyen esta segunda etapa se adentran en la exploración del extraño mundo en el que vive un brujo. Castaneda se enfrentará a peligrosos rivales, como doña Catalina, en *The Second Ring of Power* (1977); reflexiona sobre la forma en que un brujo debe afrontar el momento de abandonar este mundo – *The Eagle's Gift* (1981) – o alecciona sobre posturas y movimientos que permiten incrementar la energía vital y que han sido transmitidos a lo largo de generaciones de “hombres de conocimiento” a sus discípulos – *Magical Passes* (1998). Castaneda desarrolla algunos aspectos que ya había tratado anteriormente, como la forma de dominar los sueños – *The Art of Dreaming* (1993) – lo que permite al soñador explorar nuevos mundos, a medio camino entre las fantasías oníricas y la vida real. Su maestro, don Juan, todavía volverá a aparecer de forma esporádica en episodios-*flashback* y una vez más recorrerán los desiertos de Sonora, manteniendo los diálogos socráticos que llenaban los primeros volúmenes.

En conjunto, todos ellos constituyen el equivalente a una guía para quienes desean seguir el “camino del guerrero” y se complementan con los seminarios y cursos impartidos por Castaneda, convertido ahora en un gurú espiritual. Los libros que integran estas dos etapas son:

Tabla 5

Carlos Castaneda y su obra	
Corpus A	Corpus B
<i>The Teachings of Don Juan: A Yaqui Way of Knowledge</i> (1968)	
<i>A Separate Reality: Further Conversations with Don Juan</i> (1971)	
<i>Journey to Ixtlan</i> (1972)	
<i>Tales of Power</i> (1974)	
	<i>The Second Ring of Power</i> (1977)
	<i>The Eagle's Gift</i> (1981)
	<i>The Fire from Within</i> (1984)
	<i>The Power of Silence: Further Lessons of Don Juan</i> (1987)
	<i>The Art of Dreaming</i> (1993)
	<i>Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico</i> (1998)
	<i>The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe</i> (1998)
	<i>The Active Side of Infinity</i> (1998)

Este segundo grupo de obras presenta un estilo que difiere del primero en contenido, estructura y función del narrador. La narración no solo se aleja del relato pretendidamente objetivo que caracterizaba los primeros libros, sino que desarrolla ideas y conceptos que resultan extraños a la realidad descrita por el autor anteriormente. El resultado desde el punto de vista del lector es de alienación; el protagonista sigue siendo el mismo, como lo siguen siendo las referencias. Los lugares, los personajes y los elementos filosóficos siguen ahí, pero la construcción del mundo en que se desenvuelve la historia ha sufrido una transformación significativa. Los personajes hablan y se comportan de forma diferente y esos conceptos con los que el lector se había familiarizado durante los diálogos entre Don Juan y Carlos Castaneda aparecen

distorsionados hasta el punto de que se vuelven apenas reconocibles. La recepción de este segundo grupo de libros por parte del público lector fue de distanciamiento, que se reflejó en un descenso en las ventas y que acercó al autor a la sección de libros de “autoayuda” y “pseudomisticismo”.

Es esta acusada disonancia narrativa entre lo que hemos llamado Corpus A (los cuatro primeros libros) y Corpus B (los ocho posteriores) la que sugiere la posibilidad de que en la confección de este último haya intervenido un autor diferente. El grado de implicación de este supuesto segundo autor en la elaboración del Corpus B es, a priori, difícil de determinar, pero la convicción de que tal vez un análisis estilométrico del conjunto de la obra pueda explicar la distancia estilística entre ambos corpus es lo que motiva el presente estudio. En esencia, se trata de verificar hasta qué punto las sensaciones subjetivas producidas por un texto determinado encuentran una correlación con los elementos lingüísticos que lo integran y que son susceptibles de medición y evaluación estadística.

3.2. Una biografía llena de sombras

De hacer caso a las declaraciones del propio Castaneda a la revista *Time* (1973), Carlos Aranha Castaneda habría nacido el 25 de diciembre de 1935, en Juqueri, Brasil y habría sido criado y educado por su abuelo Oswaldo Aranha, embajador de Brasil en los Estados Unidos y posteriormente ministro de Asuntos Exteriores en su país, dado que su padre habría muerto cuando él contaba con seis años de edad y su madre ni siquiera habría alcanzado la mayoría de edad (Hughes, 1973). Sin embargo, indagaciones posteriores sugieren como fecha y lugar de nacimiento más probables, el 25 de diciembre de 1925, en la ciudad de Cajamarca, Perú (de Mille, 1976; Fikes, 1993; Marshall, 2007). Sus padres, de acuerdo a su partida de nacimiento (Fikes, 1993), se llamaban Cesar Arana Burungaray y su madre

Susana Castañeda Novoa. Su nombre completo, según los registros municipales de la ciudad de Cajamarca, fue Carlos César Arana Castañeda⁴⁴. Para Fikes, el hecho de que sus padres no estuviesen casados en el momento de concebirle, ni en el de su nacimiento, ocasionó en él un sentimiento de vergüenza que le llevó a crear un pasado ficticio y a elaborar el concepto de “*borrar el pasado*”; idea que desarrolló, más tarde, en sus libros⁴⁵.

Quienes le conocieron de niño recuerdan que siempre tuvo el deseo de emigrar a Estados Unidos y algún tiempo después de fallecer su madre, en 1949, se decidió a abandonar su tierra natal. Según los datos que Castaneda facilitó en el formulario para obtener la nacionalidad estadounidense⁴⁶, llegó a la ciudad de San Francisco en septiembre de 1951, en el *SS Yavari*, una embarcación de larga historia pero que, al parecer, nunca navegó fuera de las aguas del lago Titicaca (Fikes, 1993).

Tras cuatro años, de los cuales no ha quedado registro ni testimonio alguno, se matricula en Los Angeles City College y entre 1955 y 1959 orienta sus estudios hacia la psicología, a la vez que asiste a diversos cursos sobre escritura creativa y periodismo. A finales de 1955 conoce a Margaret Runyan, con quien comparte un gran interés por todo tipo de cuestiones metafísicas (de Mille, 1976: 26). Al año siguiente, en una escapada a Tijuana, México, se casan y firma el certificado de matrimonio con el nombre de Carlos Aranha Castaneda. La convivencia, sin embargo, solo dura seis meses, a pesar de que el matrimonio no se disuelve legalmente hasta el 17 de diciembre de 1973.

⁴⁴ De acuerdo a de Mille (1976: 26), sus apellidos revelan la mezcla de nacionalidades de sus ancestros. De Mille atribuye un origen castellano a los apellidos *Arana* y *Castañeda*, vasco a *Burungaray* y portugués a *Novoa*. En realidad, *Arana* (*ciruela*) proviene también del País Vasco, mientras que el uso del apellido *Castañeda* se distribuye por diversas provincias españolas.

⁴⁵ Castaneda llevó este precepto hasta sus últimas consecuencias, llegando a mentirle incluso a su esposa, Margaret Runyan, a quien aseguró que había nacido en 1931, en Italia, de Susana Novoa, joven estudiante de un prestigioso colegio suizo, que habría quedado embarazada tras mantener relaciones con uno de sus profesores (Runyan, 1996).

⁴⁶ Parece que es a partir de su llegada a los Estados Unidos de América, cuando Carlos Cesar Arana Castañeda modifica su segundo apellido para adaptarlo a la grafía anglosajona y pasa a escribirlo como *Castaneda*.

Castaneda entra en la Universidad de California (UCLA) en 1959 y se gradúa en antropología en 1962. No se le conocen ensayos académicos, salvo la disertación impartida en una conferencia sobre antropología, en 1968 y obtiene su doctorado en 1971. Posteriormente imparte un seminario para graduados titulado “The Phenomenology of Shamanism”, en la primavera de 1972, al que asisten medio centenar de alumnos, atraídos principalmente por el prestigio que le otorgan sus dos libros publicados hasta ese momento⁴⁷.

Michael Korda, el editor de Simon & Schuster, con quien Castaneda y su agente negociaban la publicación de sus libros, recuerda la atractiva personalidad del escritor.

“Not only was he a good talker in a town where good talkers are a dime a dozen, but far rarer, a good listener. He transformed listening into a physical act, his dark eyes fixed on me, his mobile, expressive face showing, like a good actor’s, a combination of attention, sympathy, and warm amusement. Chunky and solid as he was – he was no beauty – Castaneda had an actor’s physical grace and an exact sense of timing, together with the ability to convey, by small gestures and changes of expression, a whole range of emotion.” (Korda en Patterson, 2008: 15-16).

Castaneda era de estatura baja, para los estándares norteamericanos; entre 1,68 y 1,70 metros de estatura y de complexión fuerte, alrededor de 70 kilos de peso, de acuerdo a su ex esposa Margaret Runyan. Su pelo era negro, rizado y abundante, y su tez oscura. Acostumbraba a vestir tanto ropa deportiva como trajes de tres piezas y, ocasionalmente, sombrero. Apenas se conservan fotos del escritor, pues debido a sus hábitos reservados no le gustaba que le fotografiasen, reserva que se extendía a su rutina cotidiana. Durante su noviazgo con Runyan, ni siquiera ella sabía dónde vivía Castaneda

⁴⁷ *The Teachings of Don Juan: A Yaqui Way of Knowledge* (1968) y *A Separate Reality: Further Conversations with Don Juan* (1971).

y a pesar de que le escribía cartas románticas, las misivas llegaban sin firmar. “Carlos ya era elusivo antes de que aprendiese nada de don Juan”, asegura Runyan (de Mille, 1976: 28).

Sin embargo, Carlos Castaneda no solo era elusivo, sino también contradictorio y a la escasa información que proporcionaba sobre sí mismo hay que añadir el hecho de que esta no siempre era veraz, como hemos visto, lo que contribuye a enmarañar cualquier intento de establecer unos hechos sólidos sobre su persona. Castaneda, supuestamente, habría aprendido italiano y portugués en la granja de sus abuelos, aunque no hay constancia de que nadie le haya escuchado hablar en estos idiomas. También supuestamente, habría estudiado en una escuela privada de Buenos Aires, hasta que, debido a su carácter indisciplinado, su familia le envió a Los Ángeles para completar su educación con una familia adoptiva y habría estudiado, además, escultura en la Academia de Bellas Artes de Milán, Italia. Todos estos datos forman parte de la historia sobre sí mismo que le contó a su esposa Margaret, así como la historia de su supuesto matrimonio con una muchacha gitana y la de su período de servicio en el ejército de los Estados Unidos, en “la guerra de España” (¿?). En cualquier caso, en los registros del Departamento de Defensa no consta ningún dato sobre Carlos Castaneda (de Mille, 1976).

Las historias sobre sí mismo no siempre estaban bien elaboradas y en ocasiones resultaban tan inverosímiles que movían a la desconfianza incluso a los periodistas mejor predispuestos, como sucedió con Bruce Cook, quien tras entrevistar al escritor para la revista *National Observer* quedó tan frustrado que pospuso la publicación de su reportaje durante todo un año: “No creía que Castaneda fuese quien él decía que era”, aseguraba Cook (1973: 33).

En la entrevista que concedió a Sandra Burton, para la revista *Time*, en 1973, Castaneda afirma que intentar verificar su vida por medio de datos y fechas equivaldría a utilizar la ciencia para validar la hechicería; “le robaría la

magia al mundo”. En cualquier caso, y poco después, el escritor suprimiría sus escasas apariciones públicas y dejaría de conceder entrevistas. Castaneda se retira entonces a una casa en Los Ángeles junto a tres seguidoras o discípulas, a quien él llama “sus brujas”: Son Regina Thal, Maryann Simko y Kathleen Pohlman. Todas ellas reciben de Castaneda nombres nuevos, de acuerdo a sus nuevas vidas e identidades; Thal, pasa a llamarse Florinda Donner-Grau; Simko se convierte en Taisha Abelar y Pohlman en Carol Tiggs. Más tarde, el grupo se ampliaría con la llegada de Amalia Márquez (alias Talia Bey) y Patricia Partin (Nuri Alexander, en su “nueva existencia”). Todas ellas cortaron su relación con familias y amigos tan pronto como se unieron al círculo más próximo a Castaneda, ahora en su papel de gurú o líder espiritual⁴⁸.

Durante los años de su “retiro”, publica cinco libros más⁴⁹, aunque a partir de *Tales of Power* (1974), la figura de don Juan desaparece y las historias se convierten progresivamente en relatos místicos, a medio camino entre la narración fantástica y la autoayuda espiritual, e incluye en ellos, como personajes, a las mujeres de su círculo: Florinda, Taisha y Carol, quienes, además, le sirven de reclamo para captar nuevas adeptas en los cursos semi-privados que imparte. Además, Castaneda acostumbraba a pedir a quienes se incorporaban a su círculo más próximo que abandonasen su trabajo y mantener relaciones íntimas con él formaba parte del proceso de iniciación⁵⁰ (Wallace, 2003; Patterson, 2008).

⁴⁸ Estos elementos que implican la ruptura con la vida que hasta entonces han llevado los adeptos – familia, amigos, adopción de nombres nuevos – forma parte, por así decirlo, del manual de gestión empleado por sectas de todo tipo. La idea es aislar a los individuos de cualquier apoyo externo y crear una fuerte dependencia emocional del líder y de la estructura del “grupo”. Otros rasgos característicos son la dependencia económica y el sometimiento sexual al líder, rasgos que, como veremos, se dan también en el círculo formado alrededor de Carlos Castaneda.

⁴⁹ *Tales of Power* (1974); *The Second Ring of Power* (1977); *The Eagle’s Gift* (1981); *The Fire from Within* (1984); *The Power of Silence* (1987).

⁵⁰ Una de las adeptas de Castaneda fue Amy Wallace, hija del escritor Irving Wallace y escritora ella misma, quien a mediados de los años setenta, cuando todavía tenía diecisiete años, entró en el círculo de Castaneda y se convirtió en una de sus amantes. Wallace habla de todo ello en el libro *Sorcerer’s Apprentice* (2003).

En la década de 1990, Carlos Castaneda “reaparece” para promocionar sus enseñanzas, condensadas en una disciplina que recibe el nombre de *Tensegrity*⁵¹, una serie de ejercicios que tienen como finalidad “estirar los meridianos del cuerpo, sus canales energéticos” (Patterson, 2008: 49). Se trataría de secuencias de movimientos de origen tolteca que don Juan le habría enseñado a Castaneda y que combinan pases mágicos y ejercicios respiratorios. Estos ejercicios, de acuerdo a lo que actualmente transmiten los discípulos de Castaneda, tendrían una antigüedad de “entre 7.000 y 10.000 años” (Cleargreen, 2013):

“Don Juan explained to his students that those shamans discovered through practices that he could not fathom, that it is possible for human beings to perceive energy directly as it flows in the universe. In other words, those shamans maintained, according to Don Juan, that any one of us can do away, for a moment, with our system of turning energy inflow into sensory data pertinent to the kind of organism that we are. Turning the inflow of energy into sensory data creates, shamans affirm, a system of interpretation that turns the flowing energy of the universe into the world of everyday life that we know.” (Cleargreen, 2013).

Merece la pena mencionar que en ninguno de los libros en los que aparece don Juan se hace referencia a estas secuencias de movimientos, en los que algunos observadores han detectado ciertas similitudes con el kung fu, el karate y el tai chi (Patterson, 2008: 48). Castaneda habría estudiado durante años con el maestro de artes marciales Howard Y. Lee⁵², a quien dedica el libro *The Fire from Within*, en agradecimiento “por ayudarme a restaurar mi energía, y por enseñarme un camino alternativo hacia la plenitud y el bienestar” (1984: v).

⁵¹ El concepto de *tensegridad* – *tensegrity* – se debe al arquitecto estadounidense Buckminster Fuller y es una palabra compuesta que resulta de la combinación de *tensional* e *integrity*, que da nombre a aquellos diseños en los que varios elementos forman una estructura en equilibrio estable (Gómez Jáuregui, 2007).

⁵² Howard Y. Lee se presenta a sí mismo, hoy en día, como “maestro de la energía” y “sanador de sanadores”, por su “milagrosa habilidad para curar con la energía” y para “transferir instantáneamente a los demás la capacidad de curar”. Dirige en la actualidad *The Light of Life Institute*, desde Santa Mónica, California.

Durante los años siguientes Castaneda se dedicó a impartir seminarios ayudado por sus discípulas y colaboradoras – Thal, Simko y Pohlman, principalmente – no solo en California, sino también en Sudamérica y Europa, a precios que alcanzaban los 1.200 dólares por un fin de semana y que podían atraer a varios cientos de alumnos interesados en conocer de primera mano las enseñanzas del “brujo” (Marshall, 2007).

En 1995 fue creada la corporación Cleargreen Incorporated, con el propósito de gestionar los ingresos producidos por los cursos impartidos por Castaneda y sus discípulas, y por los libros y vídeos de enseñanza con las técnicas del sistema *Tensegrity* (Cleargreen, 2013). Aunque sus últimos libros se habían alejado de los primeros, tanto en contenido como en la forma, la preocupación por hacer frente a la muerte como un guerrero seguía muy presente en todos ellos.

“In the '90s, Castaneda told his followers that, like don Juan, he wouldn't die — he'd burn from within, turn into a ball of light, and ascend to the heavens.” (Marshall, 2007).

Hacia el verano de 1997 a Castaneda le fue diagnosticado un cáncer de hígado. La noticia se mantuvo en el más estricto secreto, dado que se suponía que un brujo como era él no podía enfermar. Sin embargo, las mujeres de su círculo más inmediato comenzaron a buscar tratamientos alternativos a medida que la salud del escritor declinaba (Marshall, 2007). A pesar de sus precauciones, quienes se encontraban cerca de ellos percibían que el ambiente entre los “iniciados” era cada vez más opresivo y las referencias al suicidio más frecuentes. A principios de abril de 1998 los más íntimos al maestro fueron vistos recogiendo diversas pertenencias y disponiéndose a cerrar la residencia del grupo y el 27 de ese mismo mes, Carlos Castaneda falleció.

A partir de ese momento, las mujeres que él llamaba sus “brujas” – Regina Thal, Maryann Simko, Kathleen Pohlman, Amalia Márquez y Patricia Partin – desaparecen. Unos pocos días más tarde, el coche de Partin es localizado en Panamin Dunes, una desértica región del Valle de la Muerte. El fallecimiento de Castaneda tarda dos meses en hacerse público y en aparecer en la prensa (Applebome, 1998; Moehringer, 1998; Valenzuela, 1998). Aunque para entonces pocos antropólogos o especialistas en cultura india conservaban alguna duda al respecto de la fantasía que contenían sus libros, las notas necrológicas mantienen una cierta reserva y eluden referirse al tema abiertamente. A día de hoy, las mujeres más próximas a él siguen ilocalizables. El abogado Richard Jennings⁵³, que durante algunos años formó parte del grupo cercano a Castaneda, cree que se quitaron la vida. Jennings considera que todas ellas lo tenían planeado de antemano, para cuando Castaneda falleciese (Marshall, 2007). Dado que la mayoría de ellas habían cortado los lazos con sus familias, nadie puso denuncia por desaparición. Tan solo el hermano de Márquez lo hizo, pero no consiguió que la policía abriese una investigación hasta dos años más tarde, cuando fueron encontrados los restos de un cadáver en el desierto y los análisis de ADN confirmaron que se trataba de los restos de Patricia Partin.

Sin embargo, Debora Drooz, albacea del testamento de Carlos Castaneda no cree que las mujeres hayan fallecido. El escritor legó la mayor parte de sus bienes a una entidad que él había creado, Eagle’s Trust, y cierta cantidad de dinero a las desaparecidas. Por su parte, su ex mujer, Margaret Runyan y su hijo Carlton Jeremy Castaneda reclamaron la parte que según ellos les correspondía⁵⁴. Al parecer, Castaneda firmó el testamento apenas

⁵³ Richard Jennings estuvo involucrado con Castaneda y su grupo entre 1995 y 1998 y realizó diversos trabajos de representación legal para ellos. Tras el fallecimiento del escritor creó la página web *Sustained Action* dedicada a recopilar información sobre Castaneda, sus seguidores y sus enseñanzas, “A website devoted to exploring and evaluating the claims of Carlos Castaneda”.

⁵⁴ Según afirma Margaret Runyan, C.J. Castaneda no es el hijo biológico del escritor, pero este lo reconoció como hijo suyo en su partida de nacimiento y contribuyó a su educación, aunque en los últimos tiempos no habían tenido relación (Applebome, 1998). C. J. Castaneda asegura que las mujeres del

cuatro días antes de fallecer⁵⁵, por lo que Carlton Jeremy sospechaba que no lo hizo por su propia voluntad, a lo que Drooz respondió que el escritor firmó los documentos ante un notario y que estaba completamente lúcido en el momento de hacerlo. Margaret Runyan falleció en enero de 2012, a los 90 años (Los Angeles Times, 2012).

3.3. Los elementos de una impostura

Ya hemos dicho más arriba que la obra de Castaneda fue saludada en un primer momento por la crítica especializada como un valioso documento antropológico y con gran entusiasmo por parte de los lectores. Posteriormente, y a pesar de las dudas suscitadas acerca de la veracidad de lo narrado, sus libros se han convertido en un referente popular. Robert Marshall asegura que su lista de admiradores incluía a John Lennon, William Burroughs, Federico Fellini y Jim Morrison (Marshall, 2007). De acuerdo a la revista *Time* (1975, 18 de agosto), los miembros del grupo de rock californiano *Eagles* se pusieron este nombre inspirados por los libros de Castaneda. La publicidad de su editorial proclama que ha vendido ocho millones de ejemplares y que su obra ha sido traducida a 17 idiomas (Amazon, web). En *Los Angeles Times*, Castaneda llegó a ser considerado como el padrino de la *New Age* (Chickering, 1988) y Robert Hughes escribió en *Times* “El encuentro entre Castaneda y don Juan parece uno de los más afortunados de la historia de la literatura, desde que Boswell fue presentado al Dr. Johnson” (Marshall, 2007).

entorno de Castaneda habrían estado impidiendo que contestase a sus llamadas o respondiese a sus cartas (Patterson, 2008).

⁵⁵ La firma del mismo tuvo lugar el 23 de abril y en el mismo acto se constituyó la fundación Eagle's Trust, beneficiaria de la mayor parte de sus propiedades. El fallecimiento se produjo a las tres de la madrugada del día 27 y el certificado de defunción señala como causa de la muerte “encefalopatía metabólica”, provocada por dos semanas de fallos hepáticos y diez meses de cáncer de hígado (Applebome, 1998b). El cuerpo fue incinerado a las pocas horas y sus cenizas trasladadas a un lugar que no fue hecho público.

No cabe duda de que la entusiasta recepción por parte de los estamentos académicos tuvo mucho que ver en la elaboración de un aura de respetabilidad que perduró en el tiempo, a pesar de las críticas que se le hicieron más tarde. En su momento, el antropólogo Paul Riesman se refirió a los tres primeros libros asegurando que “Castaneda deja claro que las enseñanzas de don Juan nos dicen algo al respecto de cómo es realmente el mundo” (Riesman, 1972: 10-12). Peter Furst, Director Asociado del Centro Latino Americano de UCLA organizó, en 1967, una serie de conferencias sobre la utilización ritual de alucinógenos en las que Castaneda tomó parte como ponente, junto al propio Furst. Ambos hablaron de las proezas físicas que habían presenciado realizar a ciertos chamanes sobre las resbaladizas rocas de algunas cataratas mejicanas. Furst se refería al chaman Ramón Medina, sobre quién había estado realizando estudios culturales y Castaneda a los saltos sobrehumanos que había visto ejecutar a don Genaro, el brujo amigo de don Juan. Furst, que describe en detalle estos sucesos en su libro *Flesh of the Gods* (1972), se mostró muy sorprendido por la similitud entre las observaciones de Castaneda y las suyas propias (de Mille, 1976).

De Mille proporciona algunas razones para lo que, visto desde la distancia, parece una sorprendente credulidad académica. Por un lado, asegura, para muchos estudiosos de la cultura indígena de los Estados Unidos, las experiencias relatadas por Castaneda suponían una “*excelente ilustración literaria del funcionamiento de la mente india*”, validando teorías y trabajos propios, que hasta entonces carecían de un soporte empírico. Además, Castaneda, al reconocer que se había involucrado personalmente en los rituales de los que daba cuenta – y consumido él mismo los hongos alucinógenos – estaba rompiendo una barrera que ninguno de sus predecesores se había atrevido a franquear o a admitir que había franqueado. De Mille pone como ejemplo al antropólogo Michael Harner, quien reconocería posteriormente que había bebido *Banisteriopsis*, una sustancia alucinógena, durante sus investigaciones con los indios jíbaros, tiempo atrás,

en 1961, y que la experiencia había resultado fundamental para comprender la cultura sometida a estudio: “I realized that anthropologists, including myself, had profoundly underestimate the importance of the drug in affecting native ideology”. (Harner, 1973). De Mille (1976) considera que algunos científicos expresaban su agradecimiento hacia Castaneda, citando su trabajo en sus propios estudios.

Este “agradecimiento” les llevaba también, en ocasiones, a reconocer un cierto valor a su obra, a pesar de que debiese ser catalogada como ficción y no como estudios científicos, pues “había servido de estímulo para considerar alternativas textuales dentro de la tradición etnográfica” (Marcus y Fisher, 1986: 40). Para otros incluso, aunque sus andanzas no fuesen más que ficción, los conocimientos filosóficos transmitidos merecían la pena: “This fake is bringing us genuine teachings!” (Staniford, 1980: 151).

Lo cierto es que los primeros libros de Carlos Castaneda llegaron en un contexto cultural propicio para recibirlos.

“In 1968, the year when Castaneda’s first book became an instant best-seller, the psychedelic movement was in full bloom. The formation of an audience ready and willing to be awed by the antics of Castaneda’s exotic mentor, don Juan Matus, took fifteen years of super-sensational media coverage of psychedelic superstars and their establishment opponents.” (Fikes, 1993: 13).

Tal y como asegura Fikes, el momento era adecuado. Había una fuerte demanda, por parte de los movimientos contraculturales, de estudios que respaldasen aquella nueva forma de entender la existencia y la relación del individuo con el mundo, una visión que intentaba imponerse y contrarrestar la enorme influencia del modo de vida tradicional. No había muchos académicos que estuviesen en la posición o con la disposición adecuada para ofrecer este

respaldo; el prestigio del doctor Timothy Leary había disminuido considerablemente y, aparte de los estudios de Mircea Eliade (1964) sobre las técnicas empleadas por los chamanes para alcanzar el éxtasis por medio de hongos y hierbas alucinógenas, el número de especialistas en este campo era escaso.

Por otro lado, la técnica narrativa utilizada por Castaneda en sus primeros libros sin duda contribuye a crear una sensación de proximidad y verosimilitud en el lector. El empleo del narrador en primera persona permite al lector identificarse desde el primer momento con el protagonista, quien a pesar de adoptar un pretendido distanciamiento, se implica, de hecho, en los sucesos que relata. Carlos Castaneda se presenta, además, como un investigador ingenuo y a la vez escéptico; dispuesto a entrar en el juego que su maestro le propone, pero al mismo tiempo incapaz de aceptar acontecimientos que chocan con una visión racional de la realidad. Su escepticismo está siempre presente y el escritor lleva esta argucia literaria hasta el último extremo, pues ante sucesos que desafían la lógica o la razón o que podrían ser atribuidos a una percepción distorsionada debida a las sustancias ingeridas, se adelanta a las dudas del lector, haciéndolas suyas y expresándolas antes de que lo haga este.

El abundante uso de diálogo entre Carlos Castaneda – el aprendiz – y don Juan – el maestro – constituye un recurso retórico que se remonta a los diálogos socráticos de Platón y cuya finalidad es la de transmitir una propuesta filosófica de forma que resulte amena para el lector. Además, se ajusta, en este caso, a la identidad de quien la transmite, un indio yaqui carente de educación formal y que no dispone, en teoría, de la destreza verbal necesaria para una tarea semejante de ninguna otra forma que no sea mediante el intercambio de preguntas y respuestas.

Hasta cierto punto su técnica narrativa se aproxima al género de no-ficción, tal y como fue formulado en su día por los escritores del *New Journalism*. En la década de 1960 y 1970 floreció en los Estados Unidos una corriente de escritura que modificaba los paradigmas narrativos que los periodistas habían venido aplicando hasta entonces. Quienes lo pusieron en práctica consideraban que podían utilizar los mismos recursos de estilo que empleaban los escritores de ficción para hacer más atractivos a los lectores sus textos. Sus reportajes dejaron de primar la objetividad por encima de todo y comenzaron a incluir una gran diversidad de puntos de vista, subjetivos, pero que proporcionaban valiosa información acerca de cómo percibían los acontecimientos quienes los habían vivido. El propio periodista dejó de ser una figura invisible y se convirtió en uno de sus propios personajes, proporcionando un relato preciso de sus sentimientos y emociones.

Entre los nombres más destacados de este movimiento, están los de Norman Mailer, Gay Talese o Tom Wolfe, quienes produjeron su obra más conocida en el mismo momento histórico en que lo hizo Carlos Castaneda. Fue Tom Wolfe quien estableció, en su libro *The New Journalism* (1973), las reglas básicas del estilo⁵⁶: El primero de ellos es la estructuración de la historia a partir de escenas concretas, trasladando al lector allí dónde se producen los hechos de forma que los presencie, por así decirlo, en primera persona. En este aspecto, los cuatro primeros libros de Castaneda se ajustan a la perfección a esta primera norma, ya que el lector acompaña al narrador y lo sitúa en el mismo lugar donde transcurre la acción. De hecho, con frecuencia el escritor interrumpe la narración de forma abrupta tras un acontecimiento excepcional o tras una afirmación memorable de don Juan. Una vez alcanzado

⁵⁶ La novela de no-ficción es una extensión de los ya de por sí largos reportajes periodísticos característicos del *New Journalism*. El ejemplo más significativo de este género es la obra de Capote, *In Cold Blood* (1966), pero otros trabajos en esa misma década – *The Armies of the Night* (1968), de Norman Mailer; *The Kingdom and the Power* (1969), de Gay Talese o *The Electric Cool-Aid Acid Test* (1968), de Tom Wolfe – están considerados igualmente excelentes muestras de esta hibridación de géneros.

el clímax de la escena, esta concluye creando un efecto dramático muy eficaz, propio de la novela de ficción.

El segundo elemento que caracteriza a este tipo de escritura es la inclusión de diálogo real, el cual, además de proporcionar inmediatez a la narración, permite construir la personalidad de los individuos retratados. Esto requiere una considerable destreza por parte del escritor, dado que tiene que dejar constancia de forma precisa de lo que dicen los personajes⁵⁷. En el caso de Castaneda, su capacidad para registrar por escrito largas conversaciones mientras estas estaban teniendo lugar, sin necesidad de interrumpirlas causa admiración, como poco. Richard de Mille considera que solo mediante las llamadas técnicas de escritura automática se podría llevar a cabo una hazaña semejante.

“Carlos’s volumes of scribbled notes should without fail be deposited in some scholarly library, not merely to show that the interviews actually occurred, but also to illustrate a remarkable recording technique. If Castaneda could demonstrate the technique in person that would be even better, methodologically.” (De Mille, 1976: 48).

En cualquier caso, y dado que la única información sobre don Juan de que dispone el lector proviene de su intervención en los diálogos, estos resultan vitales para la construcción del personaje.

Por otro lado, los otros dos elementos característicos del *New Journalism* o de la novela de no-ficción mencionados por Wolfe, esto es, el cambio frecuente de puntos de vista y la descripción de elementos que denoten la clase social de los personajes, se encuentran ausentes en los libros

⁵⁷ Truman Capote consideraba fundamental para el escritor de no-ficción esta habilidad, que él mismo aseguraba haber desarrollado tras ejercitarse durante años (Plimpton, 1966). Por su parte, Tom Wolfe, en su polémico reportaje “Radical Chic. That Party at Lenny’s” (1970) recoge fielmente los diálogos que escuchó durante la fiesta, en casa del compositor y director Leonard Bernstein, hasta el punto de que fue acusado de haber utilizado una grabadora, lo que él se tomó como un cumplido (McKeen, 1995: 77).

de Castaneda, quien nunca abandona su papel de narrador en primera persona y no considera que una descripción detallada de las ropas que visten o el lugar en el que viven los protagonistas sea relevante para la historia.

Castaneda utiliza, por tanto, algunas de las técnicas procedentes de la narrativa de ficción que fueron utilizadas por los escritores de no-ficción. Sin embargo, en su caso el número de quienes consideraban que su obra encajaba mejor dentro de la categoría de novela de ficción fue creciendo con el paso del tiempo⁵⁸. La escritora Joyce Carol Oates expresó sus dudas a este respecto en varias ocasiones. La primera de ellas en respuesta a la reseña sobre los dos primeros libros de Castaneda que *New York Times Book Review* publicaba bajo la firma de Paul Riesman y que se menciona más arriba. Oates, en una carta al periódico, reconocía el interés que despertaban en el lector los libros de Castaneda, pero se preguntaba si no se trataría de ficción pura (Oates, 1972). Algún tiempo después, en el ensayo “Don Juan’s Last Laugh”, aseguraba:

“Perhaps it takes a writer of fiction to intuit the work of a fellow artist; at any rate it seems to me beyond a doubt that this series of books is art, not merely reportorial observation.” (Oates, 1974: 10).

Y no fue la única en expresar preocupaciones similares. También lo hizo el investigador Robert Gordon Wasson, quien estudiaba y catalogaba

⁵⁸ En cualquier caso, resulta curioso comprobar que una de las críticas que con más frecuencia recibía el *New Journalism* era su falta de rigor. Con frecuencia se acusó al género de “crear” personajes literarios a partir de otros reales, de quienes se tomaban rasgos diversos para formar un único individuo que se hacía pasar por verídico (Thompson, 1987). Este fue el caso, por ejemplo, de la prostituta “Redpants” que Gail Sheehy “creo” para un reportaje en el semanario *New York* y que se repitió con “Tribal Rites of New Saturday Night”, de Nik Cohn (1976), que sirvió más tarde de base para la película “Saturday Night Fever” (Carmody, 2008). La figura de “don Juan”, de los libros de Carlos Castaneda bien podría haber sido “construida” de esta forma, al menos esto es la opinión de Jay Courtney Fikes (1993), Amy Wallace (2003) y su ex mujer Margaret Runyan (1996): “‘Don Juan’ was real. He was a real Indian, somebody Carlos actually was making trips to see. It’s just that once Castaneda got to the point of putting it all down in a readable form, the don Juan of his books became a different creature, a broad omniscient construction made up of equal parts real Indian, pure Castaneda imagination, library research and dozens of conversations and experiences with people like C.J., myself, Mike Harner, colleagues at UCLA, his grandfather and others.” (Patterson, 2008: 119).

desde hacía años las plantas que con fines medicinales y rituales utilizaban los indios mexicanos. Wasson sintió, en un primer momento, gran interés por los textos publicados por Castaneda, alguien que, como él, se ocupaba de un campo que hasta entonces había recibido muy poca atención académica. Sin embargo, albergaba ciertas reticencias, en especial relacionadas con la ausencia de información botánica y zoológica precisa. Wasson escribió a Castaneda expresándole sus inquietudes y este le remitió doce páginas con sus anotaciones de campo fotocopiadas. Se trataba de una “segunda generación” de notas, arregladas y ordenadas, no las notas primarias tomadas sobre el terreno. Estas notas recogen, en español, las conversaciones mantenidas entre Castaneda y don Juan los días 8 y 15 de abril de 1962, tal y como aparecen, traducidas al inglés en su primer libro *The Teachings of Don Juan*. En el año siguiente Wasson y Castaneda se reunieron en dos ocasiones y la opinión del primero sobre el segundo fue muy positiva: “He was obviously an honest and serious young man.” (de Mille, 1976: 47).

Sin embargo, los siguientes libros terminaron por cambiar esa primera opinión favorable del etno-botánico. Para Wasson, Castaneda había abandonado la ciencia para dedicarse a escribir romance. Sus libros no eran sino una extensa alegoría, en el mejor de los casos (Wasson, 1972; 1974).

Por su parte, el también etno-botánico Weston La Barre⁵⁹ consideraba que el ensayo que constituía la segunda parte del primer volumen de Castaneda, *A Structural Analysis*, era tan solo “un tedioso intento de poner en práctica juegos ‘levystrausianos’, que no pueden satisfacer ni a un comité, ni al público lector en general” (La Barre, 1976: 41). Este ensayo, que La Barre

⁵⁹ Weston La Barre, una de las mayores autoridades en el uso ritual de peyote, en Norteamérica, recibió el encargo de *New York Times Book Review* para elaborar una reseña crítica sobre los tres primeros libros de Castaneda, sin embargo, su opinión sobre una obra que consideraba “vulgar y pseudo-etnografía”, no resultó del agrado del editor Roger Jellinek, quien trasladó el encargo al antropólogo Paul Riesman. Más tarde, Jellinek justificó el cambio alegando que La Barre había elaborado una crítica furibunda y que le parecía que Castaneda merecía una aproximación más personal. Jellinek reconoce que esta es la única ocasión, en los años que ejerció como editor de *NYTBR*, en la que una reseña se hubiese encargado en dos ocasiones (Marshall, 2007).

consideraba “inadecuado”, “aburrido” e “intelectualmente kitsch”, tampoco satisfizo a la antropóloga y durante algún tiempo compañera de Castaneda, Barbara G. Myerhoff (de Mille 1980), a Oates (1972), o ni siquiera a Riesman (de Mille, 1976).

Por otro lado, el aspecto místico de sus libros ha sido cuestionado por Agehananda Bharati, para quien el consumo de sustancias estupefacientes no puede, en ningún caso, corroborar premisa o doctrinal o teológica alguna.

“My personal disapprobation of Castaneda’s writings stems less from their anthropological inadequacy, which is patent, than from their mystical pretentions (...) I suspect Castaneda is a fast, perceptive, highly pragmatic reader of relevant mystical texts, albeit in translation. Information sufficient for his authorly needs could be found in the Upanishads and half a dozen other translated Eastern canonical texts, or in Herbert V. Guenter’s dry *Tibetan Teutonica*, in my *Tantric Tradition*, or in Anagarika Govinda’s *Foundations* or *White Clouds*. Much of it can even be found in the sensational writings of Blavatsky, Ouspensky, Gurdjeff, and David-Neel. Castaneda’s genius, I believe, lies in transforming such materials into a pseudo-Mexican pseudo-mystical adventure tale for naïve modern readers”. (Bharati, 1980: 148-149).

En cuanto a las posibles influencias detectadas en las enseñanzas que don Juan le transmite a su discípulo Carlos Castaneda, los orígenes son muy diversos, de Mille encuentra que algunos de los “viajes” ocasionados por la ingestión de plantas alucinógenas coinciden con otros descritos por Aldous Huxley, mientras que la descripción del aura que rodea a las personas, como “un huevo luminoso”, procede de las que el escritor orientalista William Walker Atkinson (1862 – 1932) hace en sus libros. En cierta ocasión, al ser preguntado por esta similitud, Castaneda la justificó alegando que la brujería practicada por los indios norteamericanos, podría tener un origen asiático,

dado que al parecer estos llegaron al continente americano a través del Estrecho de Bering, en épocas remotas (1972).

Sin duda de entre las muchas voces críticas que se han alzado contra la obra de Carlos Castaneda destaca la de Richard de Mille por ser la que con más fuerza se ha dejado oír. Su libro *Castaneda's Journey. The Power and the Allegory*, publicado en 1976, inició una revisión rigurosa de la obra de Castaneda cuando nadie hasta entonces había cuestionado públicamente el trabajo del escritor. Su exhaustivo estudio ha llevado paulatinamente al convencimiento de que tanto el personaje de don Juan, como los acontecimientos narrados deben ser considerados exclusivamente como ficción. William W. Kelly, director del Departamento de Antropología de la Universidad de Yale, asegura que todo antropólogo de su generación contempla hoy en día a don Juan como un personaje ficticio creado por un hábil farsante. "Perhaps to many it is an amusing footnote to the gullibility of naïve scholars, although to me it remains a disturbing and unforgivable breach of ethics." (Marshall, 2007).

En *Castaneda's Journey*, de Mille somete los primeros libros a un escrutinio minucioso, detectando inconsistencias relevantes y a las que merece la pena dedicar atención. El primer aspecto que de Mille resalta es la disparidad cronológica existente entre el primer y tercer libro. Castaneda, en su pretendido registro de los encuentros mantenidos con don Juan, proporciona fechas, que, una vez ordenadas, entran en conflicto entre sí.

"Numerous metaphisicians have told us that in the nonordinary reality, in the nagual, or in some absolute realm time can stand perfectly still, can bring past and future into an etenal present, or can run in any direction. One theory of subatomic particles holds that matter going forward in time equals antimatter going backward. I would certainly not dismiss these propositions. In fact I am devoted of them. All the same, when we try to determine whether a book is a novel or a factual

report, we do not probe its subatomic structure or visit the writer in our dreams. We examine the text, judging it by commonsense rules of time and tide, clock and calendar. Marked anachronisms or logical conflicts in Castaneda's work must argue that this text is an imaginative fabrication rather than a factual report." (de Mille, 1976: 166-177).

De Mille elabora un listado de 22 experimentos con drogas de diversos tipos que se extienden a lo largo de los tres primeros libros y de 34 eventos no-ordinarios, entendidos como tales aquellos que desafían la lógica o la razón, como saltar hasta lo alto de una montaña, volar convertido en cuervo u observar la vegetación agitarse cuando no hay viento. Castaneda comienza su tercer libro, *Journey to Ixtlan* (1972), asegurando que tras reconsiderar todo su aprendizaje hasta ese momento y reexaminar sus notas de campo en las que ha recogido sus conversaciones y experiencias vividas con don Juan, ha llegado a la conclusión de que el consumo de plantas alucinógenas no es la única vía de acceso a esa realidad alternativa que don Juan le ha estado mostrando. El escritor se embarca entonces en un extenso *flash-back*, vuelve atrás en el tiempo y relata sucesos relevantes, a los que, por alguna razón, hasta ese momento no había dado importancia. Es al contrastar la situación cronológica de estos eventos con aquellos otros narrados en el primer libro cuando comienzan a aparecer los errores e incongruencias temporales, de hecho algunos de dichos eventos revisten tal importancia que hacen que de Mille se pregunte cómo es posible que no fuesen mencionados en el primer o segundo libros. A partir del contraste entre el primer y tercer libros, "una cosa es segura", afirma de Mille: "al menos uno de los dos libros, debe ser ficción" (1976: 168).

Richard de Mille aventura una hipótesis para este cambio de actitud al respecto del consumo de drogas: Quienes conocieron a Carlos Castaneda aseguran que mostraba una fuerte aversión a su consumo, ahora bien, hacia mediados de la década de 1960 la tendencia entre los movimientos

contraculturales se inclinaba a favor del uso de drogas para alcanzar un nivel superior de conciencia. El primer libro de Castaneda aprovechó este contexto social, como hemos dicho más arriba, para ganar el favor de los lectores, sin embargo, cuando en la década siguiente, las drogas perdieron gran parte de su atractivo, Castaneda comenzó a “reconsiderar” la importancia de las mismas dentro del sistema de creencias transmitido por don Juan⁶⁰. Esto hizo necesario revisar todo lo dicho anteriormente y encontrar elementos que sostuviesen todo el entramado ideológico, prescindiendo del consumo de plantas. En cualquier caso, entre el segundo y tercer libro los sucesos se acumulan de tal forma que resulta difícil creer que pudieran tener lugar.

“In the flush of first success, Castaneda had been grinding out magical events faster than the earth was turning, and his readers (and presumably his agent and his Publisher) were already howling for more. *A Separate Reality* had carried us all the way to October 1970, narrative time. *Ixtlan* would be published just two years later, calendar time. That left a two-year credibility gap in which somebody had to experience 17 of the events listed in Table 2 (as well as the events of May 1971, not listed) and Castaneda had to write a book about them, saving out at least six months and more likely a year for the publisher’s manufacturing and premarketing. To make matters worse, during the first of those two years Castaneda was actively enrolled as student at UCLA, presumably doing visible things on campus. In the spring of the second year he was lecturing at UC Irvine. Even for a sorcerer that’s a pretty tough schedule.” (1976: 172).

Por otro lado, las historias de Castaneda incluyen rituales o ejercicios, que, sin pertenecer a la categoría de sobrenaturales, parecen por sí mismos casi imposibles de poner en práctica. Un ejemplo es el episodio en el que,

⁶⁰ Resulta interesante observar que, en esas mismas fechas, el escritor Ken Kesey, que durante algún tiempo lideró desde California el movimiento a favor del consumo de LSD como medio de “trascender la realidad”, hacia finales de la década había cambiado de opinión e intentó transmitir la idea de que había llegado el momento de dejar a un lado el LSD, de evolucionar y encontrar la forma de acceder a esos “estados superiores de conciencia” sin el uso de drogas (Wolfe, 1968).

siguiendo las indicaciones de don Juan, Carlos Castaneda captura dos lagartos del desierto con intención de convertirlos en “ayudantes mágicos”. Para ello debe, entre otras operaciones, coserles los párpados y la boca utilizando una espina de cactus a modo de aguja y fibras vegetales como hilo (1968). El propio Castaneda en su relato reconoce la dificultad que ello le supuso, algo que a de Mille le parece que roza lo increíble⁶¹.

“You who have passed a steel needle and a polished tread in a bright light through a piece of inert but paper-thin leather about as wide as a newborn baby’s little fingernail without ripping it will, I am sure, appreciate Carlos’s skill in passing a perforated cholla thorn trailing an agave fiber in the darkling twilight through the tiny fragile blinking membranes that shield a living lizard’s eyes”. (1976: 42).

El uso del idioma es otra cuestión que de Mille analiza en su estudio. Supuestamente las conversaciones mantenidas entre don Juan y Carlos Castaneda transcurrieron en español, lengua que ambos dominaban. Dado que los libros fueron publicados en inglés, Castaneda introduce, en ocasiones, la traducción de ciertas frases al español. Para de Mille, esto se justifica, en un reportaje antropológico, en aquellos casos en los que la frase en el idioma original aporta un contenido semántico, un matiz, que es difícil de obtener al traducirla. Sin embargo, tras examinar todas las frases en español incluidas en sus libros y el contexto donde aparecen, llega a la conclusión de que son por completo redundantes, y por tanto innecesarias. Su finalidad solo puede ser, añade de Mille, la de reforzar en el lector – de un lector que desconozca este idioma y no perciba la mencionada redundancia – la ilusión de que la conversación original tuvo lugar en español (1976).

⁶¹ De Mille decidió poner a prueba personalmente los materiales mencionados por Castaneda para llevar a cabo una tarea semejante: una espina de cholla (*Cylindropuntia fulgida*) y fibras de agave (*Agave*). Al parecer, las fibras de agave, una vez secas, podrían hacer las veces de hilo, pero no de aguja, pues resultan endebles para esa tarea. Las espinas de cholla, por otra parte, son tan finas que se parten a la más mínima presión y nunca podrían servir para realizar perforaciones. En conjunto, resultan materiales inadecuados para una tarea como esta, “que habría sido ejecutada, además, en la penumbra” (De Mille, 1980: 442).

A este respecto ya hemos hablado de la extrañeza que produjo en el botánico Robert Wasson el que un trabajo de campo que recoge experiencias de varios años en los que se manipulan plantas y animales típicos de la región, tan solo aparezcan los nombres de unos pocos, menos de media docena y en ningún caso los términos empleados para referirse a ellos por parte de don Juan. Tampoco se incluyen muestras de la lengua natal del maestro de Castaneda, un indio yaqui que presumiblemente habría recurrido con frecuencia a expresiones, nombres o palabras de su propia cultura para explicar unas creencias de siglos de antigüedad. Tras nueve años de trato frecuente, Carlos Castaneda tan solo habría aprendido dos palabras yaquis: *yori* (hombre blanco) y *Torim*, el nombre de un asentamiento yaqui. El narrador asiste a diversos encuentros de don Juan con otros indios yaquis en los que parece razonable pensar que se emplearía alguna expresión en su lengua nativa. Sin embargo, y a pesar de que en alguna ocasión Castaneda afirma que recoge las conversaciones con una grabadora oculta, no hay referencias a este idioma.

Esta ausencia de datos y evidencias materiales se extiende al conjunto del estudio de Castaneda. No existen fotos, grabaciones, ni testimonios sobre la existencia de don Juan o de don Genaro, más allá de lo que el escritor incluye en sus libros. Sus notas de campo, que se supone extensas, nunca han sido vistas por nadie, con la excepción de las doce fotocopias remitidas a Robert Wasson, a pesar del valor que a este tipo de documentos se le atribuye dentro de los estudios antropológicos.

Con tales deficiencias metodológicas, parece razonable preguntarse cómo pudieron pasar los libros de Carlos Castaneda por trabajos antropológicos serios durante tanto tiempo, e incluso ser publicado el primero de ellos – *The Teachings of Don Juan* – por la editorial de UCLA. Parte de la respuesta a esta pregunta bien podría estar en los nombres del académico que

dirigió su trabajo de tesis, Harold Garfinkel, y del que avaló la publicación este primer volumen, Walter Goldschmidt.

Garfinkel, profesor de sociología de la Universidad de California, es el creador, así mismo, de la disciplina llamada *etnometodología*, de acuerdo a la cual toda realidad puede considerarse subjetiva, o al menos “intersubjetiva”, dado que se construye a través del consenso logrado entre dos hablantes. El objetivo de la misma es establecer los métodos y las prácticas a través de los cuales los miembros de una sociedad dan sentido al mundo en el que viven (Garfinkel, 1967). Garfinkel había dirigido y supervisado la tesis de Castaneda – la cual, a través de las enseñanzas de don Juan, ejemplifica las propuestas del propio Garfinkel – y “le habría hecho reescribirla hasta tres veces” (de Mille, 1976: 80), dando como resultado el diario de las conversaciones entre Castaneda y don Juan y el ensayo que lo acompaña, *A Structural Analysis*. Una vez finalizado, y con el visto bueno de Garfinkel, y de los antropólogos Clement Meighan y Robert B. Edgerton, el manuscrito llegó a la editorial de la Universidad de California (ibid., 1980). En las páginas iniciales, Castaneda agradece a estos tres académicos su ayuda en la elaboración del estudio. Walter Goldschmidt, por entonces responsable del Departamento de Antropología escribe el Prefacio que comienza con la afirmación de que “el libro es a la vez etnografía y alegoría” y termina asegurando que el autor ha demostrado, en este trabajo, “la destreza esencial de la buena etnografía – la capacidad de introducirse en un mundo extraño”(Goldschmidt, [1968] 1998: xxiii-xxvi).

No cabe duda de que el beneplácito de nombres consagrados académicamente tuvo algo que ver en la consideración que recibieron sus textos a partir de ese momento. De Mille se pregunta sobre las razones por las que nadie expresó en público su disconformidad con esta situación y se responde a sí mismo, aludiendo a que, como otros profesionales, los antropólogos prefieren solventar estas cuestiones en privado. Además, añade,

140

ficción o no, los libros de Castaneda ilustran un ideal del trabajo de campo que muchos antropólogos intentan transmitir a sus alumnos, quienes, por otro lado, “están ansiosos por recibir historias excitantes sobre las materias que estudian” (ibid., 1976: 83).

El sociólogo y antropólogo Stephen O. Murray coincide con de Mille en su convicción de que la academia es, en general, reacia a exponer públicamente a quienes realizan trabajos poco serios.

“There is normally no reason to pillory suspect work in public. Such work is not cited, not built upon, not considered in reviews of literature, and not enshrined as established knowledge in textbooks. Because there is a flood of publication, much of it unimportant, scientists avoid dwelling on what they consider suspect and concentrate on keeping abreast of what their peers consider significant.” (Murray, 1980: 198).

A pesar de todo, hacia mediados de la década de 1970, ciertos especialistas, como el antropólogo Michael Harner, todavía consideraban que sus libros, independientemente de la credibilidad con que fuesen leídos, seguían siendo válidos. De este modo afirma: “I think Castaneda’s work is 110 per cent valid. He conveys a deep truth, though his specific details can often be justifiable questioned.” (Harner, 1978: 45).

Algunos se sentían conmovidos a un nivel emocional por su relación personal con el escritor, como en el caso del profesor de antropología y psiquiatría de UCLA, Douglas Price-Williams:

“I had intense experiences with Carlos that are difficult to explain. You see, you can’t say his work is factual, but you can’t say it’s false either. It’s so much more complex than that. He did have profound experiences on his own. And he had a great deal of ethnographic knowledge. He also engaged in elaborate role playing that he took to

the point that I think he could no longer tell the difference. But the thing that set Carlos apart was his genius for talking all this and communicating it in a way that really moved people.” (Patterson, 2008: 114-115).

Para otros, como Stephen J. Reno, profesor de Historia de las Religiones, en la Universidad de Leicester, “a pesar de su rechazo a una interpretación literal, la leyenda de don Juan sigue teniendo un enorme atractivo” (Reno, 1980: 258) y para el escritor William Irwin Thompson la obra de Castaneda responde a una necesidad perentoria de la sociedad de “integrar la ciencia y lo oculto” (Thompson en de Mille, 1980: 105). La también escritora Amy Wallace resume los sentimientos ambivalentes que continúa despertando en quienes le conocieron y en quienes leyeron sus libros:

“Carlos had begun as a genuine seeker, a true philosopher, he had ended as a tyrant watching over a cult of terrified followers. Power had wielded its legendary seductions, illness had weakened him terribly. But nothing, I believe, can subtract from the sincerity and beauty of his early Works. To take their wisdom and leave the rest would be to take the best of Castaneda. ... I do not believe that Carlos was a con man, who callously sought money and women. I think that he believed in his dream to the last, and did his best to make it come true; that he made terrible mistakes as the years went by, due to poor judgement, narcissism, and illness; and in his last decade he did create an abusive cult. This is no black and white tale, for Carlos was not a shifty huckster but a misguided philosopher whose experience of power was corrupting. Thus he damaged many lives, at the same time exalting many others . . . I sometimes find it remarkable that my deep feeling for Carlos persist so strongly. Despite what I’ve learned about the perversity of a sexual guru/disciple relationship, I love him still, and I continue to grieve for him.” (Patterson, 2008: 109-110).

Desmontar una impostura es una tarea ingrata. Lo es fundamentalmente por dos motivos, el primero de los cuales radica en su dificultad metodológica. Siempre resulta mucho más fácil demostrar que algo *sucedio* que probar la afirmación contraria. Si alguien acudiese hoy a nosotros con un manuscrito en el que relatase unos acontecimientos que le sucedieron en el día de ayer – por ejemplo, que voló convertido en ave tras pronunciar un encantamiento encontrado en un viejo libro de magia – y asegurase que todo es cierto, pero sin aportar ninguna evidencia, estaría trasladándonos a nosotros *la carga de la prueba*⁶² que, en rigor, le corresponde a él. Seríamos nosotros quienes tendríamos que encontrar evidencias de que lo que afirma no sucedió realmente, algo que bien puede convertirse en una labor imposible. Es la razón por la que, en el proceso penal, quien hace una acusación debe demostrar que es cierta, ya que de hacerlo a la inversa – exigir al acusado que demuestre su inocencia – estaríamos creando una posición de insoportable indefensión.

Esta viene a ser, en resumen, la situación a que han tenido que enfrentarse Richard de Mille y quienes, como él, han intentado exponer a la luz pública el fraude que constituyen los libros de Carlos Castaneda. Por razones sobre las que de Mille solo ha conseguido especular, a Castaneda no se le exigieron en su momento pruebas que demostrasen la veracidad de un relato que resultaba a todas luces extraordinario. Una vez superado este primer filtro – el de la validación académica – el escritor se instaló en la confortable posición de quien traslada a sus críticos la obligación de encontrar evidencias de que lo que él afirmaba era falso. A pesar de ello, para lograrlo, de Mille sometió los libros de Castaneda a un minucioso análisis, en una auténtica labor de *close reading*, buscando discrepancias, incongruencias y

⁶² En principio, quien hace una afirmación, tiene la obligación de aportar las pruebas que la demuestren. De no hacerlo así estaría trasladando la responsabilidad a sus hipotéticos críticos y cometiendo la *falacia de la ignorancia*, o ausencia de evidencia en sentido contrario, lo que equivale a afirmar que algo es cierto dado que no se ha demostrado que sea falso (Michalos, 1969: 370; Joyce, 1919). El filósofo Bertrand Russell ilustró este problema con la afirmación de que “una tetera de porcelana china gira alrededor del Sol, entre la Tierra y Marte”, una proposición absurda, pero sin duda difícil de demostrar que *no es cierta* (Russell, 1952).

desajustes narrativos que demostrasen que lo que en ellos se relata no era verdad.

Quienes han leído los trabajos de Richad de Mille no albergan duda alguna de que ha logrado su propósito y nos parece conveniente aclarar que quien firma este trabajo se encuentra entre ellos. Sin embargo, y para ser honestos, debemos considerar objetivamente la validez de las *pruebas* que de Mille aporta, de entre las cuales, tal vez la de más peso sea la que señala los desajustes cronológicos entre unos libros y otros. Conviene, para ello, hacer una distinción entre *pruebas* e *indicios*: la *prueba* científica debe ser objetiva, verificable y reproducible y sustentar, sin lugar a dudas, la afirmación que se pretende demostrar. El *indicio*, por otro lado y de acuerdo a la definición que proporciona el Diccionario de la RAE, es un “fenómeno que permite conocer o inferir la existencia de otro no percibido”. La diferencia entre ambos, en la práctica, resulta más tenue, más sutil, de lo que aparece al formular su descripción teórica y en los procesos judiciales con frecuencia la consideración de uno u otro resulta crucial a la hora de emitir un veredicto. En ocasiones, la acumulación de indicios se traduce en el *convencimiento* de quien debe evaluar, aun sin llegar a constituir entre todos ellos una prueba como tal.

Y este es el caso, nos parece, con el que nos encontramos aquí. Castaneda mintió repetidamente al respecto de su origen, lugar y fecha de nacimiento, cierto. Sus libros presentan claros desajustes temporales, cierto, también. En ellos se narran, no solo hazañas extraordinarias e imposibles de acuerdo a la lógica – saltar a lo alto de una montaña, ver caer una misma hoja de un árbol repetidamente, observar el aura de una persona en forma de huevo luminoso... – sino también otras que sin serlo, resultan implausibles, como el ya mencionado episodio en el que cose los párpados a dos lagartos, o cuando caminan él y su maestro bajo el abrasador sol del desierto sin dar muestras de incomodidad. En la región donde transcurren los hechos, por ejemplo,

abundan los escorpiones, sin embargo, y a pesar de que ambos pasan allí días enteros durmiendo incluso sobre el terreno, en ningún momento se los menciona. Todo esto también es cierto, está ahí y es verificable. Ahora bien, ¿demuestran estos datos que lo que se relata en los libros es falso? Nos parece que no. Es improbable que sea cierto, pero no imposible. Son *indicios*, no *pruebas*, y por tanto sugieren y apuntan en una dirección determinada, pero no demuestran fehacientemente que don Juan no existiese, por ejemplo.

Los libros de Castaneda son, con toda probabilidad, un fraude, pero hoy tenemos el *convencimiento* de su falsedad no tanto debido a las pruebas que esforzados investigadores como de Mille o Jay C. Fikes hayan aportado, sino más bien debido a que en todo este tiempo ni el propio Castaneda en su momento, ni ningún otro investigador más tarde, han aportado pruebas *positivas* de que lo narrado en sus libros es cierto.

Por otra parte, el segundo motivo por el que deshacer el nudo de una mentira resulta poco gratificante se debe a las reacciones emocionales que genera. No parece probable que la revelación de que Castaneda fue un impostor haya sido motivo de alegría para nadie, ni tampoco que sean muchos los que se han sentido agradecidos por conocer la verdad. Si algún sentimiento han provocado los estudios que se han realizado en este sentido, ha sido de decepción, vergüenza e incluso hostilidad.

Para un número considerable de personas los libros de Carlos Castaneda supusieron la confirmación de que existe algo en este mundo que trasciende la realidad visible y cotidiana, la confirmación de que todos tenemos en nuestro interior el potencial necesario para realizar proezas increíbles, de que la magia, en suma, existe. Sus libros van dirigidos no a la parte racional del individuo, sino a aquella otra que regula las emociones y gran parte del talento de Castaneda como escritor reside en despertar en los lectores el deseo de creer que lo que están leyendo es cierto. De ahí que haya

sido preciso otra generación, como dice William W. Kelly, para revisar sus libros con el distanciamiento necesario para verlos como lo que realmente son: una fábula, una metáfora, una alegoría, que condensa ideas filosóficas ya existentes y las presenta en un nuevo contexto, pragmático y poético a la vez, capaz de emocionar a los lectores.

Otros, quienes otorgaron validez científica a lo narrado, sin duda han de sentirse un poco avergonzados de su credulidad, por haber permitido que sus emociones se interpusiesen en el camino de su razonamiento. Sin duda, algunos de ellos han de sentir una cierta aversión hacia quien les deja en mal lugar, poniendo en duda su competencia académica.

Por nuestra parte, el proceso de exponer las razones por las que esta supuesta obra antropológica debe ser considerada como ficción, resultaba necesario para la tarea que nos ocupa. Dado que pretendemos verificar si los cuatro primeros libros firmados por Carlos Castaneda fueron escritos por la misma persona que escribió los ocho últimos, es importante determinar previamente si el texto que aparece en cada uno de ellos es homogéneo o no lo es, es decir, si es una persona o son varias quienes *hablan* en sus libros.

Ya hemos dicho que los libros están compuestos en buena medida por las conversaciones sostenidas entre el narrador y otros personajes. Pues bien, si los libros fuesen el relato fiel – como pretende su autor – de situaciones en las que se vio involucrado y hechos de los que fue testigo, habría que suponer que los parlamentos de los otros personajes transcriben sus palabras tal y como las pronunciaron. Esto obligaría a expurgar los libros de todas aquellas partes en las que hablan otros que no sean el propio narrador/escritor.

Si por el contrario consideramos los libros como ficción fruto de la inventiva de su autor, Carlos Castaneda, podemos atribuirle todo lo que estos incluyan, sea quien sea el personaje que hable en un momento determinado.

Este será, por tanto, el punto de partida desde el que abordaremos la siguiente fase de nuestro trabajo, en la que también determinaremos el proceso que seguiremos para realizar el análisis estilométrico.

3.4. Aproximación estilística a la obra de Carlos Castaneda

La evaluación estilística de la obra de Carlos Castaneda presenta varios problemas, derivados en su mayor parte de su dificultosa adscripción genérica. Si, como todo parece indicar, se trata de textos cuyo contenido debe ser atribuido exclusivamente a la imaginación de su autor, parece que una primera aproximación a los mismos debería hacerse desde la perspectiva de la categoría de “ficción”.

Ahora bien, surge de manera inmediata la primera distinción con respecto a otras obras literarias igualmente ficticias y es la de que, en estas, los autores no simulan la naturaleza de la historia que están trasladando al lector. Por mucho que se esfuercen por teñirlas de verosimilitud, ningún escritor – por regla general – oculta que la esencia última de su trabajo es fruto de la fantasía.

Diversos recursos estilísticos, bien conocidos a estas alturas, contribuyen a crear en el receptor de la literatura de ficción la sensación de que los hechos narrados son tan próximos a la realidad que – tal y como Aristóteles sugería – si no han sucedido, bien podrían haberlo hecho. Incluso en ciertos casos los autores, advierten al lector que lo que van a leer a continuación “está basado en hechos reales”, trasladándole de esta forma la decisión de elegir hasta qué punto creer o no en lo que está leyendo. El contrato o acuerdo tácito que se establece entre ambos ofrece, pues, variantes, pero rara es la vez que se quiebra por ninguna de las dos partes: ni el uno

elige creer a pies juntillas lo que lee, ni el otro pretende que lo narrado sea una crónica veraz de lo acontecido.

En el caso de Castaneda esto no es así. Desde el primer momento el escritor nos propone un relato detallado de hechos supuestamente reales, en los que él mismo tomó parte. Su narración pretende ser un reportaje aséptico, avalado por el método científico y – más tarde – sancionado por el estamento académico. El lector se encuentra entonces ante un tipo de contrato literario en el que se le solicita que crea sin reparos, que suprima cualquier atisbo de escepticismo. Este tipo de relación no es inusual, pues la encontramos en los textos históricos, informes científicos y, en general, en las crónicas periodísticas.

Así pues, el lector de sus primeros libros acepta sin objeciones su contenido, por mucho que en ocasiones lo que cuenta contravenga a la razón. Para ello, además del respaldo académico ya mencionado, el escritor juega con la complicidad de una gran parte de los lectores, que se sumergen de buen grado en la realidad alternativa, en los mundos paralelos que Castaneda y su maestro don Juan proponen. Y es que el contexto es propicio para ello; durante los años sesenta y setenta se produce una eclosión de movimientos culturales que ofrecen nuevas formas de contemplar el mundo. Las filosofías importadas de Oriente desafían la lógica mecanicista que ha venido imponiéndose en Occidente desde mediados del siglo XIX e incluso ciertos estamentos científicos comienzan a dudar de que sea posible conocer la realidad más allá de la apariencia que nuestros sentidos imperfectos nos revelan. Se vuelve, además, la mirada hacia culturas indígenas, primitivas tecnológicamente, pero que conservan el aprecio por la sencillez y un contacto con la Madre Naturaleza que el hombre civilizado ha perdido. En este contexto, los escritos de un antropólogo que aseguraba haber encontrado una antiquísima tradición mesoamericana capaz de abrir las puertas a una

realidad paralela fueron saludados, si no con alborozo, sí con el agrado que se reserva para quien corrobora lo que la intuición parecía venir sugiriendo.

Este agrado, sin embargo, no fue unánime. Para algunos la violencia, por así decir, que supuso forzar a la razón a aceptar hechos que la lógica negaba resultó excesiva. La desconfianza hacia los estudios presentados por Castaneda creció, avivada por una absoluta falta de pruebas que respaldasen sus afirmaciones. Investigadores como De Mille o Fikes dedicaron esfuerzos considerables a intentar desenredar la enmarañada madeja de embustes y verdades a medias que Castaneda había ido elaborando con los años.

Una encendida polémica se extendió durante casi dos décadas, entre los partidarios del escritor y sus detractores, lo que equivale a decir entre quienes tenían fe y quienes exigían pruebas. Uno de los efectos que este debate tuvo fue el de relegar a segundo plano el valor literario de sus trabajos, pues las dos posturas enfrentadas obviaban cualquier otra consideración: Si sus relatos eran verídicos, su significado en términos culturales, antropológicos y filosóficos eclipsaban cualquier posible mérito artístico y, si no lo eran, si eran una impostura, debían ser desechadas al completo.

Sin embargo, es innegable que las andanzas del escritor, sus extrañas peripecias y la descripción del mundo que su maestro don Juan propone tuvieron un efecto emocional en muchos lectores que condicionó la forma en que recibieron, no solo los libros en sí mismos, sino también las críticas posteriores que ponían de manifiesto lo implausible de sus relatos. Castaneda como escritor fue capaz de alterar la percepción racional de muchos receptores de sus mensajes y si midiésemos el valor de una obra literaria por el impacto que esta tiene en sus lectores, habría que concluir que fue considerable. Se trata de una afirmación discutible, pues ya en su momento W. K. Wimsatt y M. Beardley expusieron en su bien conocido ensayo “The Affective Fallacy” (1949) los peligros que implica tomar como elemento de

juicio la reacción que una pieza literaria provoca en los receptores de la misma. Wimsatt y Beardley advierten de la confusión que en ocasiones se produce entre la obra y sus resultados (lo que *es* y lo que *hace*), pues que estos últimos sean notables, no conduce necesariamente a que la primera también lo sea, desde un punto de vista literario. En cualquier caso, su propia afirmación no se haya exenta de opiniones discrepantes – al igual que lo está la escuela crítica a la que ambos pertenecen, el *New Criticism* – sobre todo por parte de tendencias surgidas en la segunda mitad del pasado siglo. El escritor y crítico literario C. S. Lewis en *An Experiment in Criticism* ([1961] 1992) señala una de las consecuencias inmediatas de separar la obra en sí de la respuesta del público lector: la división de la literatura en “alta” y “baja”. Lewis, acertadamente, vincula el momento y las circunstancias en que un libro es leído con la reacción que este genera en el lector, de lo cual el caso que nos ocupa es un buen ejemplo.

Otros críticos, en cambio, son más propensos a considerar la importancia de la relación que se establece entre la pieza literaria y los lectores. Es el caso de Stanley Fish o Norman Holland, impulsores de la escuela crítica *reader-response* que estudia la reacción que despierta una pieza literaria determinada en ciertos lectores y los mecanismos psicológicos que impulsan esta respuesta. En los experimentos llevados a cabo por Holland (1975) se pone de manifiesto que es el sustrato psicológico de los lectores lo que les lleva a reaccionar de determinada manera ante un texto literario:

“Thus, I learned that people’s responses to literature involved a transformation by means of forms acting like defenses, of drives, impulses, and fantasies back and forth from the most primitive strata of psychic life to the highest. Given such a model, one could understand the social, intellectual, or moral themes people found in literature as the highest level of this dynamic and continuing process of transformation. One could explain the way readers respond to literary characters as

though they were real people, when they are patently not real and often not very realistic.” (Holland, 1975: xii).

Como ya hemos dicho, un determinado estado mental compartido por una parte importante de la sociedad y vinculado a los movimientos culturales alternativos de aquellos años tuvo sin duda mucho que ver con la recepción que obtuvieron las obras de Castaneda. No cabe duda de que leídas hoy en día han de generar una respuesta distinta, pues incluso obviando toda la literatura crítica con la misma, el clima imperante de escepticismo hacia las proezas sobrenaturales influiría de forma notable en los lectores.

Conviene, sin embargo, hacer una distinción que separe en dos categorías el contenido de sus libros: por un lado están aquellos pasajes en los que se describen hechos que desafían la razón o la lógica y, por otro, los momentos en los que don Juan imparte a su discípulo lecciones sobre la vida y la forma en que esta debe ser vivida. Entre los primeros incluimos todo tipo de manipulaciones y procedimientos “mágicos”, que para el lector no iniciado en tales prácticas – como lo es el propio personaje de Castaneda – resultan incomprensibles.

En *Tales of Power*, por ejemplo, Castaneda relata un encuentro inquietante en mitad de la noche durante una excursión por el desierto mejicano, cuando cree ver una figura humana acechando en la oscuridad:

“I then experienced a strange pain in the pit of my stomach. Something seemed to rip inside me and I could not hold the muscles of my midsection tense any longer. At the very moment I let go, the dark shape of an enormous bird, or some sort of flying animal, lurched at me from the chaparral. It was as if the shape of the man had turned into the shape of a bird. I had the clear conscious perception of fear. I gasped and then let out a yell and fell on back.” (1974: 10 -11).

Don Juan se ríe de los temores de su discípulo y le proporciona una explicación que no resulta muy satisfactoria, en términos racionales.

“What happened out there, don Juan?” I finally asked

“You had an appointment with knowledge,” he said, pointing with a movement of his chin to the dark edge of the desert chaparral. “I took you there because I caught a glimpse of knowledge prowling around the house earlier. You might say that knowledge knew that you were coming and was waiting for you. Rather than meeting it here, I felt it was proper to meet it on a power spot. Then I set up a test to see if you had enough personal power to isolate it from the rest of the things around us. You did fine.”

“Wait a minute!” I protested. “I saw the silhouette of a man hiding behind a bush and then I saw a huge bird.”

“You didn’t see a man!” he said emphatically. “Neither did you see a bird. The silhouette in the bushes and what flew to us was a moth. If you want to be accurate in sorcerer’s terms, but very ridiculous in your own terms, you could say that tonight you had an appointment with a moth. Knowledge is a moth.” (1974: 11-12).

Este tipo de “explicación-que-no-explica-nada” es habitual a lo largo de toda la relación que mantienen maestro y alumno. Cada vez que un fenómeno *extraño* tiene lugar, don Juan proporciona a Castaneda información de muy escaso valor, ya que este no puede desprenderse de los procesos lógicos a los que está acostumbrado y su conocimiento del “mundo de los brujos” siempre parece demasiado escaso para darle sentido a lo que ve. En todo proceso comunicativo, emisor y receptor deben compartir una base de información mínima que permita al segundo decodificar adecuadamente el mensaje que el primero le remite, pero este no es el caso: ni Castaneda, ni los

lectores estamos familiarizados con esa otra realidad a la que don Juan se refiere constantemente. De ahí que por mucho que este intente transmitir, sus mensajes carecen de significado.

Esto se pone de manifiesto en otro ejemplo, en el que el Castaneda aprendiz, expresa al maestro sus dudas al respecto de si estará haciendo lo correcto al revelar en sus libros los conocimientos recibidos, pero este lo tranquiliza.

"But maybe I'm revealing things I shouldn't, don Juan."

"It doesn't matter what one reveals or what one keeps to oneself," he said. "Everything we do, everything we are, rests on our personal power. If we have enough of it, one word uttered to us might be sufficient to change the course of our lives. But if we don't have enough personal power, the most magnificent piece of wisdom can be revealed to us and that revelation won't make a damn bit of difference."

He then lowered his voice as if he were disclosing a confidential matter to me. "I'm going to utter perhaps the greatest piece of knowledge anyone can voice," he said. "Let me see what you can do with it."

"Do you know that at this very moment you are surrounded by eternity? And do you know that you can use that eternity, if you so desire?"

After a long pause, during which he urged me with a subtle movement of his eyes to take a statement, I said that I did not understand what he was talking about. "There! Eternity is there!" he said, pointing to the horizon. Then he pointed to the zenith. "Or there, or perhaps we can say that eternity is like this." He extended both arms to point to the east and west. We looked at each other. His eyes held a question.

"What do you say to that?" he asked, coaxing me to ponder upon his words. I did not know what to say.

"Do you know that you can extend yourself forever in any of the directions I have pointed to?" he went on. "Do you know that one moment can be eternity? This is not a riddle; it's a fact, but only if you mount that moment and use it to take the totality of yourself forever in any direction."

He stared at me. "You didn't have this knowledge before," he said, smiling. "Now you do. I have revealed it to you, but it doesn't make a bit of difference, because you don't have enough personal power to utilize my revelation. Yet if you did have enough power, my words alone would serve as the means for you to round up the totality of yourself and to get the crucial part of it out of the boundaries in which it is contained." (1974: 7- 8).

Este recurso literario, en el que uno de los personajes apela a un conocimiento hermético para justificar acontecimientos incomprensibles, funciona sorprendentemente bien a la hora de acallar las dudas del lector, a quien se transmite la idea implícita de que "si no entiende, es porque no sabe". Se trata de una idea arraigada en el ethos cultural de numerosos pueblos: el conocimiento ilumina y la ignorancia equivale a la oscuridad. Los textos sagrados abundan en ejemplos de la confusión y desconcierto que invaden a quienes carecen de un mínimo de información relevante, desde la Biblia, donde se señala que los caminos del Señor son inescrutables (Isaías, 55), hasta la tradición zen, en la que los koan con que los maestros tratan de forzar el desarrollo espiritual de sus discípulos resultan absurdos para los no-iluminados (Watts, [1957] 1999; Suzuki, 1994). O bien desconocemos los propósitos de quien rige los destinos de los seres humanos o bien carecemos de la mirada con que el iluminado contempla el mundo; en cualquier caso nuestra ignorancia nos impide encontrar sentido a ciertos hechos cotidianos o

a la explicación que proporcionan “quienes saben”. Una situación, por cierto, que nos resulta cada vez más familiar al común de los mortales, enfrentados a unos avances tecnológicos que utilizan procesos que escapan a nuestra comprensión y que nos obligan a conformarnos con aprender a “hacer que funcionen”.

Castaneda, en cambio, quiere saber los porqués, quiere encontrar una lógica que le permita entender a qué se enfrenta y en sus dudas inacabables y desconfianza perpetua encontramos el segundo recurso que el autor utiliza para lograr la complicitad del lector. Cada vez que algo inusual sucede, cada vez que don Juan hace una proposición *extraña* para quienes no vivimos en el mundo de los *brujos*, el personaje de Carlos Castaneda adelanta sus objeciones, antes de que lo haga el lector. Castaneda verbaliza las dudas que surgen en la mente del lector y al hacerlo así, este se identifica con aquel:

"What's it like to see, don Juan?"

"You have to learn to see in order to know that. I can't tell you."

"Is it a secret I shouldn't know?"

"No. It's just that I can't describe it."

"Why?"

"It wouldn't make sense to you."

"Try me, don Juan. Maybe it'll make sense to me."

"No. You must do it yourself. Once you learn, you can see every single thing in the world in a different way."

"Then, don Juan, you don't see the world in the usual way any more."

"I see both ways. When I want to look at the world I see it the way you do. Then when I want to see it I look at it the way I know and I perceive it in a different way."

"Do things look consistently the same every time you see them?"

"Things don't change. You change your way of looking, that's all"

[. . .]

"But you won't be able to recognize anything, since nothing is the same; so what's the advantage of learning to see?"

"You can tell things apart. You can see them for what they really are."

"Don't I see things as they really are?"

"No. Your eyes have learned only to look. Take, for example, the three people you encountered, the three Mexicans. You have described them in detail, and even told me what clothes they wore. And that only proved to me that you didn't see them at all. If you were capable of seeing you would have known on the spot that they were not people."

"They were not people? What were they?"

"They were not people, that's all."

"But that's impossible. They were just like you and me."

"No, they were not. I'm sure of it." I asked him if they were ghosts, spirits, or the souls of dead people. His reply was that he did not know what ghosts, spirits, and souls were. I translated for him the Webster's

New World Dictionary definition of the word ghosts: "The supposed disembodied spirit of a dead person, conceived of as appearing to the living as a pale, shadowy apparition." And then the definition of spirit: "A supernatural being, especially one thought of... as a ghost, or as inhabiting a certain region, being of a certain (good or evil) character."

He said they could perhaps be called spirits, although the definition I had read was not quite adequate to describe them.

"Are they guardians of some sort?" I asked.

"No. They don't guard anything."

"Are they overseers? Are they watching over us?"

"They are forces, neither good nor bad, just forces that a brujo learns to harness."

"Are they the allies, don Juan?"

"Yes, they are the allies of a man of knowledge." (1971: 25-27).

Castaneda mantendrá su escepticismo hasta el final, incluso después de tomar parte en numerosos eventos extraordinarios y de los esfuerzos de su maestro por hacerle entender las leyes que rigen este mundo que se superpone al cotidiano. Al concluir su aprendizaje aparece casi tan ignorante como al comenzarlo. Cuando acude al encuentro con otros aprendices de don Juan – de los cuales lo desconoce prácticamente todo – todos parecen saber más que él:

“The witches don’t know that the Nagual told us everything. They think that they are the only ones who know. It took two Toltecs to make us. We are the children of both. Those witches...”

“Wait, wait, Pablito,” I said, putting my hand over his mouth.

He stood up, apparently frightened by my sudden movement.

“What do you mean that it took two Toltecs to make us?”

“The Nagual told us that we are Toltecs. All of us are Toltecs. He said that a Toltec is he receiver and holder of mysteries. The Nagual and Genaro are Toltecs. They gave us their special luminosity and their mysteries. We received their mysteries and now we hold them.”

His usage of the word Toltec baffled me. I was familiar only with its anthropological meaning. In that context, it always refers to a culture of Nahuatl-speaking people in central and southern Mexico which was already extinct at the time of the Conquest.

“Why did he call us Toltecs?” I asked, not knowing what else to say.

“Because that’s what we are. Instead of saying that we are sorcerers or witches, he said that we are Toltecs.” (1977: 180-181).

Al parecer, don Juan había transmitido estas enseñanzas a todos, menos a Castaneda. Este, a pesar de haber ocupado su lugar y haberse convertido en el nuevo *nagual* – el líder del grupo – precisa que le expliquen qué es lo que está sucediendo. El recurso más utilizado por el narrador es el intercambio pregunta-respuesta entre él y quien quiera que se encuentre en la escena en ese momento, ya sea su maestro, los discípulos de este o cualquier otro personaje. Castaneda pregunta sin cesar y el lector reconstruye la historia que aquel intenta transmitir a través de las respuestas que recibe. Sin embargo, si

al principio en los primeros libros esto permitía una cierta identificación con el protagonista, el abuso de este recurso hace que su ignorancia termine por parecer impostada, un truco narrativo para poner al día al lector de lo que subyace tras lo que el protagonista ve, oye y experimenta.

En cuanto a la segunda categoría en la que clasificábamos el contenido de los libros de Castaneda y que mencionábamos algunas páginas atrás, nos parece que es en ella donde residen los elementos que provocan la respuesta emocional de los lectores. A lo largo de las cuatro primeras obras, el viejo indio yaqui traza un mapa espiritual del mundo que su discípulo creía conocer tan bien y establece las reglas estrictas que guían la vida de un *brujo* o, como él prefiere llamarlo, un *hombre de conocimiento*. Don Juan describe una realidad despiadada, llena de peligros y amenazas, tanto físicas como emocionales y se refiere al mundo oculto de los *brujos*, pero se trata de una descripción de un mundo que el lector reconoce como propio, como si de una metáfora precisa y acertada se tratase. Don Juan habla de humildad y orgullo a la vez, como elementos que guían al *hombre de conocimiento*; humildad al reconocer su propia pequeñez y transitoriedad, y orgullo por vivir su vida con coraje y determinación:

“The beetle emerged from a deep hole and stopped a few inches away from my face. It seemed to look at me and for a moment I felt that it became aware of my presence, perhaps as I was aware of the presence of my death. I experienced a shiver. The beetle and I were not that different after all.

Death, like a shadow, was stalking both of us from behind the boulder. I had an extraordinary moment of elation. The beetle and I were on a par. Neither of us was better than the other. Our death made us equal. My elation and joy were so overwhelming that I began to weep. Don Juan was right. He had always been right. I was living in a most

mysterious world and, like everyone else, I was a most mysterious being, and yet I was no more important than a beetle.” (1972: 190).

El *hombre de conocimiento* es ante todo un guerrero, que temple su ánimo y vive de forma impecable, siempre a la caza de *poder personal*, siempre consciente de su muerte inevitable:

"Personal power is a feeling," he said. "Something like being lucky. Or one may call it a mood. Personal power is something that one acquires regardless of one's origin. I already have told you that a warrior is a hunter of power, and that I am teaching you how to hunt and store it. The difficulty with you, which is the difficulty with all of us, is to be convinced.

You need to believe that personal power can be used and that it is possible to store it, but you haven't been convinced so far." (1972: 127).

En estos cuatro primeros libros, el narrador es capaz de crear escenas emotivas, que conjugan la dura aspereza del mundo de los *brujos*, con una suerte de equilibrio cósmico que otorga a cada ser humano la posibilidad de acceder a una realización superior. En cierta ocasión, mientras caminan por una ciudad mexicana, Castaneda y don Juan contemplan a un grupo de niños de la calle, limpiando los zapatos de los clientes de un restaurante y alimentándose de las sobras de comida:

"Do you feel sorry for them?" don Juan exclaimed in a questioning tone.

"I certainly do," I said.

"Why?"

"Because I'm concerned with the well-being of my fellow men. Those are children and their world is ugly and cheap."

"Wait! Wait! How can you say that their world is ugly and cheap?" don Juan said, mocking my statement. "You think that you're better off, don't you?"

I said I did; and he asked me why; and I told him that in comparison to those children's world mine was infinitely more varied and rich in experiences and in opportunities for personal satisfaction and development. Don Juan's laughter was friendly and genuine. He said that I was not careful with what I was saying, that I had no way of knowing about the richness and the opportunities in the world of those children. I thought don Juan was being stubborn. I really thought he was taking the opposite view just to annoy me. I sincerely believed that those children did not have the slightest chance for any intellectual growth.

I argued my point for a while longer and then don Juan asked me bluntly, "Didn't you once tell me that in your opinion man's greatest accomplishment was to become a man of knowledge?"

I had said that, and I repeated again that in my opinion to become a man of knowledge was one of the greatest intellectual accomplishments.

"Do you think that your very rich world would ever help you to become a man of knowledge?" don Juan asked with slight sarcasm.

I did not answer and he then worded the same question in a different manner, a thing I always do to him when I think he does not understand.

"In other words," he said, smiling broadly, obviously aware that I was cognizant of his ploy, "can your freedom and opportunities help you to become a man of knowledge?"

"No!" I said emphatically.

"Then how could you feel sorry for those children?" he said seriously.

"Any of them could become a man of knowledge. All the men of knowledge I know were kids like those you saw eating leftovers and licking the tables." (1971: 14-15).

Es en momentos como este que los textos alcanzan sus más altas cotas emotivas, cuando don Juan se refiere a aspectos de la vida familiares y comprensibles para el lector. Tal vez el hecho de que con frecuencia sus afirmaciones escapan a nuestro entendimiento provoca que, cuando somos capaces de atrapar el significado de sus mensajes, nos sintamos doblemente gratificados.

Con todo, estos pasajes se concentran en aquellos cuatro primeros libros a los que hemos denominado Corpus A, mientras que en los ocho siguientes – el Corpus B – resultan mucho más escasos. En ausencia de la poderosa figura de don Juan, el narrador se ve obligado a recurrir a argucias que le permitan seguir desgranando los conocimientos del maestro. Castaneda descubre que hay una parte de su relación con don Juan que había olvidado por completo y en determinado momento accede a esa parte de su subconsciente, gracias a ciertas maniobras corporales. O bien, en su relación con los otros discípulos de don Juan se pone de manifiesto que el maestro enseñó partes distintas de su conocimiento a cada uno de ellos, como si de las piezas de un puzzle se tratase, y entre todos deben intercambiarse la información que les falta:

"Holy Jesus! We are remembering the other self!" she exclaimed, her voice almost ordering on hysteria. Then she calmed down and went on talking in a subdued tone. "Evidently we've already been there and the only way of remembering it is the way we're doing it, by hooting off our dreaming bodies while dreaming together."

"What do you mean, shooting off our dreaming bodies?" I asked.

"You yourself have witnessed when Genaro used to shoot off his dreaming body," she said.

"It pops off like a slow bullet. It actually glues and unglues itself from the physical body with a loud crack." (1981: 91).

Las "explicaciones" se vuelven todavía más extrañas, más propias de *brujos* que cuando el propio don Juan impartía su sabiduría de primera mano y el lector va distanciándose paulatinamente de un contexto que cada vez le resulta más ajeno y en el que cada vez le es más difícil encontrar puntos de referencia familiares. Tal vez consciente de ello, a partir del segundo libro de la segunda temporada, *The Eagle's Gift* (1981), Castaneda vuelve a introducir nuevos encuentros con don Juan, situados en una época anterior a su desaparición y lo hace con un recurso poco convincente, desde un punto de vista literario.

"The organization of don Juan's instruction was predicated on the idea that man has two types of awareness. He labeled them the right side and the left side. He described the first as the state of normal awareness necessary for everyday life. The second, he said, was the mysterious side of man, the state of awareness needed to function as sorcerer and seer. Don Juan divided his instruction, accordingly, into teachings for the right side and teachings for the left side." (1984: 4).

Al parecer, cuando don Juan impartía sus conocimientos hacia el lado izquierdo de Castaneda, estos quedaban registrados en algún lugar remoto de su memoria al que no pudo acceder hasta mucho más tarde. Pero no solamente la justificación para la reincorporación de don Juan resulta extravagante, sino que las propias descripciones del maestro yaqui se han vuelto todavía más fantásticas, si cabe:

“For seers it means that there are fortyeight types of organizations on the earth, fortyeight types of clusters or structures. Organic life is one of them.”

“Does that mean that there are fortyseven types of inorganic life?”

“No, not at all. The old seers counted seven bands that produced inorganic bubbles of awareness. In other words, there are forty bands that produce bubbles without awareness; those are bands that generate only organization.” (1984: 80).

Estas descripciones son ahora demasiado irreales, las relaciones entre los personajes demasiado ajenas a la sociedad cotidiana, las explicaciones demasiado enmarañadas y la demanda que el narrador hace al lector para que suspenda su incredulidad resulta, en suma, excesiva.

En *The Power of Silence* Castaneda recurre una vez más al truco de los conocimientos olvidados, para justificar el hecho de que nunca antes ha mencionado conversaciones que ahora parecen de importancia tal como para dedicarles un libro:

"You will need a lifetime to remember the insights you've had today," he said, "because most of them were silent knowledge. A few moments from now you will have forgotten them. That's one of the unfathomable mysteries of awareness." (1987: 9).

El aspecto técnico de todo el asunto implica un golpe en cierto punto preciso del costado izquierdo que don Juan propina a Castaneda y que le provoca un cambio en su nivel de conciencia. El alumno pierde su excepcional claridad mental y olvida lo que ha sucedido en los momentos previos. De hecho buena parte del libro aborda el tema del grado de conciencia que poseen los seres humanos y de la forma en que los *brujos* logran incrementarlo mediante ciertas manipulaciones, “moviendo el punto de encaje”. Para ello utilizan dos técnicas o procedimientos: el arte del acecho y el control consciente del sueño. Y al control de los sueños dedica el siguiente volumen, *The Art of Dreaming* (1993), donde da cuenta de las técnicas empleadas por los *brujos* para manipular sus sueños usando la voluntad. También aquí aparece la explicación al hecho de que Castaneda no recordase apenas nada sobre los otros aprendices de don Juan. Su maestro solo le había permitido interactuar con ellos mientras se encontraba en estado de “segunda atención”, por lo que es mucho más tarde, al desbloquear esa parte de su memoria, cuando puede acceder a estos recuerdos reprimidos.

Esta amnesia selectiva es un recurso poco elaborado, desde un punto de vista narrativo. Sin embargo, a estas alturas, la evolución que habían ido experimentando los libros de Castaneda había cribado a su público lector dejando solo a los incondicionales, cuyo nivel de exigencia era con toda probabilidad bajo en lo que se refiere a la estructura de la historia. Sus libros se habían convertido en un producto especializado, dirigido a un sector del mercado muy concreto, aunque numeroso a pesar de todo, pues es preciso recordar que además de en el mercado anglosajón, se distribuían internacionalmente.

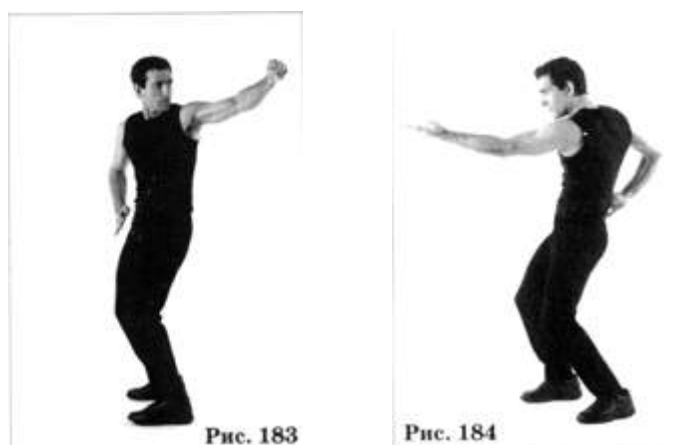
El volumen *Magical Passes. The Practical Wisdom of the Shamans of Ancient Mexico* (1998), en cambio, supone una extraña mezcla de esoterismo y realidad mundana. Dedicado a difundir los movimientos y gestos “mágicos” que Castaneda venía mostrando desde hacía algún tiempo en los seminarios

que impartía por diversas ciudades del mundo, comienza con un “disclaimer” que, a modo de encantamiento destinado a proteger al autor contra las demandas judiciales, recomienda a los lectores la conveniencia de que consulten con su médico de cabecera antes de ponerlos en práctica, para prevenir lesiones, especialmente si se trata de mujeres embarazadas.

“34. Hurling the Assemblage Point Like a Disk from the Shoulder

A deep rotation of the waist is made to the left, which propels the right arm to swing to the left side of the left leg. Then the motion of the waist, moving in the opposite direction, propels the left arm to swing to the right side of the right leg. Another motion of the waist propels the right arm to swing again to the left side of the left leg. At this point the left hand reaches back instantly with a circular motion to grab onto something as if it were solid, from the area behind the shoulder blades. (fig. 183).

The left hand takes it in a swinging circular motion to the front of the body and up to the level of the right shoulder. The palm of the clenched hand faces upward. From this position, the left hand, with a flick of the wrist, makes a hurling motion, as if to hurl forward something solid, like a disk (fig. 184).”



Está claro que se trata de un manual práctico, como su propio título indica, que muestra el desplazamiento que su narrativa había experimentado a lo largo de los años, abandonando toda pretensión académica para centrarse en un público que demandaba este tipo de literatura, relacionada con el

movimiento social y cultural que durante aquellos años recibió el nombre de *New Age*⁶³.

El último volumen, *The Active Side of Infinity*, en el que Castaneda repasa diferentes eventos de su vida en apariencia banales, pero de significado trascendente, finaliza explicando por qué no murió al saltar del risco de las montañas mexicanas, al concluir su aprendizaje:

“Yes, I had jumped into an abyss, I said to myself, and I didn't die because before I reached the bottom of that gully I let the dark sea of awareness swallow me. I surrendered to it, without fears or regrets. And that dark sea had supplied me with whatever was necessary for me not to die, but to end up in my bed in L.A.” (1998: 119).

Al despertar, Castaneda toma conciencia de que ya no es el que era; ha cambiado, está solo en un mundo misterioso, que ahora debe recorrer sin su maestro para guiarle. Lo cierto es que ese salto con el que se cierra la tetralogía que compone el Corpus A – sucediese o no – marcó un antes y un después en su vida y en su producción literaria. El elemento clave no fue tanto la hipotética muerte del narrador, como la desaparición de don Juan, puesto que es este y no Carlos Castaneda la fuerza que en realidad anima sus libros. El narrador ha construido para sí una figura anodina, sin fuerza, sin el *poder personal* que envolvía la figura del viejo indio yaqui. Su ignorancia omnipresente, su torpeza para iniciar cursos de acción que desemboquen en nuevas y emocionantes experiencias y, sobre todo, su incapacidad de generar una filosofía vital absorbente como lo hacía su maestro, a quien teóricamente ha sustituido, ocasiona que los ocho libros que siguen sean un pobre remedo de los cuatro primeros.

⁶³ Se trata de una etiqueta aplicada de forma imprecisa al conjunto de prácticas, creencias y en general filosofía de vida que se distancian de la ciencia y el positivismo racional.

A partir de aquel punto en el que don Juan desaparece, el narrador, incapaz de producir nuevos textos que mantengan la tensión emocional, se ve obligado a traer de vuelta a quien ha sido todo el tiempo el verdadero protagonista de sus historias. El problema radica en que, por alguna razón, es incapaz de devolverle la vida de forma convincente; el don Juan de la segunda parte parece un impostor, alguien que se hace pasar por el auténtico y que intenta distraer al lector con su incomprensible cháchara esotérica.

Hay todavía otra cuestión que a nuestro criterio debe ser tomada en cuenta a la hora de valorar literariamente la obra que firma Carlos Castaneda y es el propósito del autor. Se trata de un aspecto vinculado a la recepción por parte del público lector, que hemos tratado más arriba y, al igual que este, es rechazado por algunas corrientes críticas como elemento válido a la hora de enjuiciar los méritos de una pieza literaria. Si el texto en sí mismo debe ser tomado como el origen único y exclusivo del aprecio crítico, los motivos o propósitos que hayan guiado al escritor al elaborarlo deben, por consiguiente, ser considerados irrelevantes (Wimsatt y Monroe, 1946).

No cabe duda de que quienes defienden esta postura lo hacen con argumentos fundados y no nos parece que estas páginas sean el lugar adecuado para entrar, a este respecto, en un debate en profundidad. Sin embargo, estamos convencidos de que el contexto debe ser tenido en cuenta como un elemento más que contribuye al aprecio de una pieza literaria y daremos, en defensa de nuestra opinión, un único motivo. La literatura en general y una obra literaria sometida a examen en concreto no debería, en principio, recibir una consideración diferente a cualquier otro artefacto producido por el ingenio humano. La regla para determinar el valor de dichos artefactos establece que la cuestión debe resolverse respondiendo a la pregunta de si cumplen o no la función para la que han sido elaborados. La puesta en práctica de esta regla parece obvia cuando sopesamos si apreciar o no una herramienta o artilugio de cualquier clase, pero es de utilidad incluso

168

aplicándola a los actos de las personas. ¿Realiza de forma eficiente (y aquí deben incluirse consideraciones de tiempo y esfuerzo invertidos) su cometido el mecanismo que hemos diseñado? ¿Hemos conseguido la finalidad que nos llevó a iniciar una acción determinada? Si la respuesta a estas preguntas es “no”, debemos considerar fallidas nuestra creación o nuestras iniciativas, independientemente de que hayamos obtenido beneficios adicionales que no figuraban en nuestro propósito inicial. Si por el contrario la respuesta es positiva, parece razonable considerarlas como válidas, dado que han cumplido su objetivo y no sería lógico exigirles ningún mérito adicional.

Es cierto que al aplicar este criterio a la literatura nos encontramos con la incómoda situación de tener que dar por meritorias ciertas piezas que parecen carecer de valor artístico. Tomemos como ejemplo algunos éxitos literarios recientes dentro del género que se ha dado en llamar *young-adult fiction*, como son la serie *Crepúsculo* (Meyer, 2005) o *Los Juegos del Hambre* (Collins, 2008). Desde un punto de vista puramente literario habría que concluir que su valor es escaso, sin embargo, habrá que aceptar también que han proporcionado a sus autoras la popularidad y gratificación económica que sin duda buscaban cuando las escribieron. Incluso deberíamos anotar en su haber el placer que su lectura ha proporcionado a una cantidad muy considerable de lectores, pues ha de ser algo que con toda seguridad se habían propuesto tanto Meyer como Collins y que va asociado a los logros mencionados anteriormente.

Así pues, ¿debemos desechar por completo estas piezas literarias por no alcanzar los estándares estéticos establecidos por un grupo de críticos de forma arbitraria? Si lo hiciéramos, sería igual que si despreciásemos el coche de madera y cartón que un padre construye a su hijo y con el que consigue ganar la carrera popular de su barrio, solo porque no está a la altura de los vehículos que fabrican las factorías comerciales. Tanto los libros de Meyer y Collins, como el coche de juguete de nuestro ejemplo, son artefactos

refinados, elaborados de forma diestra y que realizan a la perfección la tarea para la que fueron diseñados. Ni los unos pretendían competir con escritores de la talla de Roth, DeLillo o Pynchon, ni el otro con Seat, Renault o Ford.

Escribir una pieza literaria capaz de agradar a un gran número de lectores – lo que suele denominarse un *best seller* – es una proeza que no debe en absoluto ser minimizada y cuyo mérito bien podría equipararse a quien escribe un libro que es recibido con parabienes unánimes por parte de la crítica especializada. Quienes se encuentran en el primer caso son capaces de conectar con la psique colectiva, identificar correctamente los temores, ansiedades y deseos de la población y convertir todo ello en historias que materialicen sus sueños y sean leídas con placer.

En el fondo subyace el debate sobre la existencia o no de una *alta* y una *baja* literatura, a la que Lewis se refería más arriba, pero, en nuestro caso concreto, lo expuesto en las líneas anteriores nos lleva a destacar el hecho notable de que Castaneda logró sus objetivos evidentes de forma impecable, gracias a sus libros. Obtuvo el reconocimiento académico en un primer momento y la fascinación del público más tarde; y cuando decidió convertirse él mismo en un maestro espiritual, la personificación del *nagual* literario, sus textos fueron una herramienta sumamente eficaz en la consecución de su propósito. Creemos, pues, que al menos este mérito debe serle reconocido, si bien esta propuesta suscita, a su vez, nuevas cuestiones, en esta ocasión de tipo ético y sobre el valor real de un artefacto diseñado para el engaño. Es cierto que con el paso del tiempo su aura se ha ido desvaneciendo, pero es dudoso que Castaneda pretendiese pasar a la posteridad. Si en alguna ocasión pensó en lo que quedaría de él cuando ya no estuviese presente, sin duda los versos de Juan Ramón Jiménez, que incluye en *Journey to Ixtlan*, expresan lo que sintió:

...Y yo me iré. Y se quedarán los pájaros
cantando;
y se quedará mi huerto, con su verde árbol,
y con su pozo blanco.

Todas la tardes, el cielo será azul y plácido;
y tocarán, como esta tarde están tocando,
las campanas del campanario.

Se morirán aquellos que me amaron;
y el pueblo se hará nuevo cada año;
y en el rincón aquel de mi huerto florido y encalado,
mi espíritu errará, nostálgico...

Y yo me iré; y estaré solo, sin hogar, sin árbol
verde, sin pozo blanco,
sin cielo azul y plácido...
Y se quedarán los pájaros cantando.

Poemas agrestes (1910-1911).

Castaneda creó una ficción extensa que abarcó toda su vida y parte de la cual, una parte importante, está formada por su producción literaria. Mantener esta ficción hasta su aliento final debió ser una tarea monumental, digna de la voluntad inquebrantable del guerrero que don Juan describe en sus libros. Por nuestra parte y desde la distancia solo podemos imaginar la tremenda soledad que un embuste de tal magnitud hubo de provocar.

Quien se separa voluntariamente, como Castaneda lo hizo, de aquellos que le rodean mostrando una falsa imagen y ocultando su verdadera naturaleza, habita en el mismo mundo que el resto de la gente, pero al mismo tiempo se mantiene separado de todos ellos. Es muy posible que no fuese esa su intención inicial, pero una vez adentrado por la senda de sus primeros

libros ya no le fue posible volver sobre sus pasos y deshacer lo andado. Se colocó a sí mismo en una situación paradójica: admirado por sus seguidores incondicionales y recompensado materialmente, pero a la vez alienado de toda relación humana sincera. En este sentido, Castaneda vivió una auténtica realidad alternativa y es muy posible que los sentimientos de nostalgia y soledad que esta situación le provocaron le sirviesen a menudo de inspiración para escribir los pasajes más conmovedores.

**4. Estudio de las obras atribuidas a Carlos
Castaneda: Aspectos preliminares**

4.1. Marco teórico

La atribución de autoría se basa en el hecho constatado de que cada individuo expresa una misma idea de forma diferente (Bailey, 1973; Bender, 2007; Coulthard, 2005c; Simpson, 2009). Los estudios empíricos revelan que esta diferencia es tan acusada que permite individualizar a cada autor.

“As Malcolm Coulthard has pointed out, based on the work of some computational linguists, if you highlight any ten-word string from any document, and then paste that string into a search engine (e.g., Google) or database (e.g., Lexis) window, you will find that your string is unique among the vast array of documents available. This shows enormous intra-authorial variation in the use of language. We really do not say things the same way twice. What this all means is that the challenge facing authorship attribution research is to discover ways in which the variation in the use of language between authors is demonstrably greater than the variation within a particular author and to focus on the presence or absence of these distinguishing features.” (Solan, 2013: 573).

Ahora bien, si nuestro objetivo final es determinar si un texto T1 es similar o diferente estilísticamente de otro texto T2, antes será preciso establecer con claridad *qué* es T1 y *qué* T2, o lo que es lo mismo, obtener aquellos rasgos o elementos que les proporcionan identidad a ambos, diferenciada o no. La Ley de la Identidad establece que cada cosa/elemento es igual a sí misma/mismo (Russell, [1912] 2009) y esto es así a pesar de las transformaciones que esta/este cosa/elemento experimente en el devenir temporal, lo que significa que está formada por elementos/unidades que

cambian, pero también por elementos/unidades que permanecen y a estos últimos los llamamos *esencia* (Michalos, 1969; Stanford, 2012). No cabe duda de que algo en mi hijo, por ejemplo, me permite reconocerlo como tal a pesar de los cambios que experimenta a lo largo del tiempo. Es este algo, esta esencia que permanece inmutable, aquella que le otorga su identidad y la que permite diferenciarlo de otros de su misma especie. Encontrar esta esencia es, pues, indispensable para poder discriminar entre diferentes elementos, por similares que sean entre sí.

Por otra parte, los estudios de atribución de autoría parten de la hipótesis de que dos cosas/elementos con el mismo origen, aun siendo diferentes entre sí comparten en su *esencia* ciertos rasgos que los vinculan de forma directa a su procedencia común⁶⁴ (Blasco y Ruíz, 2009; Coulthard y Johnson, 2007; Ellegard, 1962). De esta premisa se sigue que cada uno de los textos producidos por un escritor concreto, aun cuando presenten un aspecto externo distinto, portan en sí mismos ciertos atributos que los identifican como obras del mismo autor.

En consecuencia, la tarea de quien pretenda atribuir un texto a un autor concreto precisa de un primer paso que consiste en determinar los elementos que se hallan presentes en todos los escritos producidos por dicho autor. Debemos obtener aquellos rasgos estilométricos que aparecen de forma regular en un texto, en cualquier parte del mismo que elijamos y en cualquier pieza textual producida por un mismo autor, incluso aunque el contexto temático y temporal varíe. Si al dividir un mismo texto dubitado en tres partes encontrásemos que cada una de ella posee elementos o cualidades por completo diferentes, ¿Qué parte es la que compararíamos con el elemento indubitado? ¿Cómo atribuiremos un texto dubitado a un escritor, si al tomar

⁶⁴ Ya veíamos en los primeros capítulos que este razonamiento sustenta toda la ciencia forense, en sus diferentes campos de estudio. El ejemplo más evidente lo tenemos en el investigador policial que pretende atribuir una serie de delitos a un mismo autor, para lo cual precisa establecer un vínculo común entre todos ellos y el sospechoso, por ejemplo, las huellas dejadas por una misma herramienta, por una misma arma, o, incluso, la presencia de un mismo *modus operandi* en todos los casos.

tres obras indubitadas de dicho escritor, por ejemplo, cada una de ellas aparece diferente de las demás? ¿Deberemos hacer la comparación con la primera, con la segunda o con la tercera? De ahí la importancia de que los rasgos de identidad que elijamos se hallen siempre presentes y en una proporción estable.

4.2. Antecedentes metodológicos

El caso que nos ocupa, la determinación de si ciertas obras literarias han sido escritas o no por un autor concreto, presenta serias dificultades prácticas. Como ya hemos visto anteriormente, el método de atribución de autoría habitual se basa en la proximidad entre los valores que arroja el análisis de las obras dubitadas y aquellos otros procedentes de la obra u obras indubitadas (Abbasi y Chen, 2005; Baayen, Van Halteran, Neijt y Tweedie, 2002; Grant, 2008; McMenamin, 2002). Es decir, la autoría se atribuye a aquel de entre los autores candidatos cuyos textos más se aproximen al texto dubitado.

Ahora bien, para que este método funcione, se necesita como condición *sine qua non* un panel cerrado de candidatos, entre los cuales se tenga la certeza de que se encuentra el autor del texto dubitado. Cuando esto no es así no hay forma de tener la seguridad de que el candidato a quien se atribuye la autoría sea realmente el autor de texto.

Por otro lado, en situaciones similares a la que aquí se propone en las que tan solo se dispone de un autor candidato, no es posible poner en práctica el citado método de la proximidad, pues, ¿qué cercanía entre el texto dubitado y el indubitado deberíamos considerar suficiente como para atribuirlo al único autor candidato? Esta era una de las objeciones que esgrimían quienes

disputaban los resultados del estudio de autoría del *Libro de Mormón*, que veíamos anteriormente (Hilton, [1982] 1997; Midgley, 1997).

Patrick Juola resuelve este problema creando artificialmente un panel de candidatos, tal y como le hemos visto hacer en el caso “Rowling / Galbraith” (Juola, 2013a) y más tarde ante los tribunales, en la resolución de una petición de asilo político (Juola, 2013b). Este sistema, sin embargo, presenta, en nuestra opinión, deficiencias metodológicas que lo invalidan, pues al no tener la certeza de que el autor del texto se encuentra entre los componentes de ese panel de candidatos, tampoco podemos estar seguros de que aquel a quien se lo atribuimos por proximidad sea quien estamos buscando.

Es verdad que Juola toma escritores que reúnen características similares a las del autor “sospechoso” – género literario, sexo, nacionalidad, etc. – para formar un panel de candidatos lo más homogéneo posible⁶⁵. Sin embargo, esto no es más que una solución parcial, que no invalida el hecho de que sigue siendo un panel de candidatos abierto, esto es, que *el autor buscado puede estar o no incluido* en él⁶⁶. Para formar un auténtico panel de candidatos cerrado, dentro del cual se encontrase sin lugar a dudas el autor del texto analizado, el investigador debería tomar a toda la población de autores que hayan tenido la posibilidad material de escribirlo, algo que sucede en pocas ocasiones.

⁶⁵ Se trata, por lo demás, de un procedimiento estándar, puesto en práctica en las diligencias policiales de reconocimiento fotográfico o en rueda. Los investigadores deben presentar ante el testigo de los hechos, al sospechoso acompañado de un grupo de personas del mismo sexo, edad, complexión, vestimenta y, en general, mismo aspecto físico. El testigo señalará entonces, de entre todos ellos a aquel que más se parezca al recuerdo que guarda de la persona que cometió el delito que presencié. Este es, como puede apreciarse, el mismo procedimiento utilizado por el lingüista forense cuando atribuye la autoría a aquel escritor cuyos textos más se parecen al texto dubitado.

⁶⁶ Los falsos positivos, o identificaciones erróneas por parte de testigos oculares, constituyen un problema bien conocido por los investigadores, dentro del proceso penal. Se trata de un defecto del sistema en el que se conjugan un método deficiente y unos recuerdos inexactos o contaminados (Ebbinghaus, [1885] 2013). La consecuencia es, en ocasiones, una atribución de autoría incorrecta, a pesar de lo cual se sigue poniendo en práctica regularmente. La lingüística forense, al copiar el procedimiento, se expone también a la misma posibilidad de error en la atribución.

Por nuestra parte, al plantearnos el enfoque con que abordaríamos nuestro estudio, consideramos que debíamos evitar este procedimiento por su escasa fiabilidad y, para solventar el problema de la ausencia de candidatos alternativos, diseñamos un método específico de atribución de autoría aplicable en aquellos casos en los que, como aquí, se dispone de un solo candidato.

4.3. Determinación del método de análisis empleado

Así pues, con esta idea en la mente y después de numerosas pruebas, creamos un algoritmo de atribución de autoría aplicable a situaciones de candidato único. Su fundamento se apoya en la constatación de que ciertos elementos textuales se repiten en la obra de un mismo autor con regularidad y, acumulados en número suficiente, permiten la atribución de un texto anónimo o dubitado (Argamon y Levitan, 2005; Chaski, 2005; Coutlhard, 2004; Craig, 1999; Grieve, 2007). Este sistema consta de tres etapas:

1. **Establecer los Rasgos de Identidad estilística del autor analizado (RdI).** Para ello tomaremos aquellos marcadores que proporcionen información de mayor utilidad o, en otras palabras, aquellos que posean un mayor poder discriminante, que en el presente caso serán los que menor variación presenten dentro de la producción textual objeto de estudio.
2. **Establecer el Rango de Autoría del autor analizado (RdA).** Asumiendo que incluso en aquellos Rasgos de Identidad Estilística que ofrezcan valores más estables una cierta variación es inevitable, determinaremos qué oscilación de dichos valores resulta razonable admitir en el autor indubitado.

3. **Establecer el Límite de Autoría Razonable (LAR).** Habiendo constatado que incluso los textos producidos por un autor experimentan una cierta variación con respecto a otros textos del mismo autor, es preciso determinar igualmente cuál es la desviación máxima admisible del Rango de Autoría, a partir de la cual se considera con certeza que un texto dubitado ha sido escrito por un autor diferente del candidato. Esta desviación máxima o Límite de Autoría Razonable se obtendrá a partir de un estudio estadístico realizado sobre un corpus de entrenamiento.

De este modo, en nuestro estudio podríamos seleccionar como RdI las palabras “for”, “by” y “from”, porque aparecen en todos los textos indubitados de que disponemos y además lo hacen manteniendo siempre frecuencias de aparición muy similares.

El siguiente paso será establecer el RdA para cada uno de ellos y lo haremos a partir de la variación que observamos en los textos indubitados. Si la frecuencia de aparición de “from”, pongamos por caso, oscila entre 0,5 y 0,7%, parece razonable suponer que esa misma oscilación de 0,2% se puede dar en cualquier otro texto de este autor a partir de la cifra mayor y de la cifra menor. Así el RdA del rasgo “from” estará entre 0,3 y 0,9%.

Una vez hecho esto, analizaremos el texto dubitado y comprobaremos si en él la frecuencia de aparición de “from” está entre estos dos valores (0,3 – 0,9%), en cuyo caso se dará una *concordancia*, o no lo está y tendremos una *infracción*. Como quiera que incluso en los textos producidos por un mismo autor se producirán algunas *infracciones*, debido a la variación intra-autor ya mencionada, es preciso determinar el número máximo de *infracciones* a partir del cual se puede asegurar que el texto ha sido escrito por otro autor diferente del candidato. Este valor constituirá el LAR, que se determina previamente mediante un estudio estadístico para el que se han tomado 30 obras literarias

180

de ficción, escritas por 10 autores diferentes y se han sometido a análisis aplicando el algoritmo descrito.

El desarrollo y puesta en práctica de este algoritmo se describirá en las páginas siguientes, donde detallamos el proceso seguido en el análisis de las obras estudiadas y exponemos finalmente las conclusiones a que los resultados nos han conducido.

4.4. Marco empírico

4.4.1. Objetivos del estudio

- a. **Objetivo específico:** El objetivo específico que nos hemos fijado en este trabajo consiste en determinar si alguna de las obras que integran la segunda etapa literaria del escritor Carlos Castaneda difiere del resto en grado suficiente como para asegurar que ha sido escrita por otro autor. Esta diferencia debe ser cuantificable en términos susceptibles de medición por métodos automáticos.
- b. **Objetivos generales:** Los objetivos generales que pretendemos alcanzar son, en primer lugar, llevar a cabo nuestro análisis mediante técnicas estadísticas sencillas que permitan, incluso a quienes carezcan de conocimientos específicos en esta disciplina, verificar todo el proceso. Además, pretendemos desarrollar un método fiable de atribución de autoría en los casos de candidato único, sin necesidad de elaborar artificialmente un panel de candidatos cerrado. Por último, pretendemos contribuir con los resultados de este trabajo al debate sobre el alcance y los límites de las técnicas estadísticas de atribución de autoría.

4.4.2. Fases del estudio:

Hemos dividido el estudio en fases que agrupan las distintas tareas de acuerdo a una secuencia lógica, dado que cada una de ellas resulta necesaria para la ejecución de la que sigue. Así, obtenemos la siguiente distribución:

Fase 1: Selección de los marcadores estilísticos, en la que determinamos qué marcadores vamos a utilizar en nuestro análisis, atendiendo a su fiabilidad probada en estudios previos y a su facilidad de extracción.

Fase 2: Descripción de las herramientas informáticas. Se describirán brevemente los programas informáticos utilizados, en un momento u otro, para llevar a cabo la extracción automática de los datos que necesitamos para nuestro estudio.

Fase 3: Pruebas con el Corpus de Entrenamiento, en las que aplicaremos el algoritmo diseñado para este tipo de casos con la doble finalidad de verificar su eficacia y de establecer el Límite de Autoría Razonable que necesitaremos posteriormente.

Fase 4: Análisis de las obras de Carlos Castaneda. En esta etapa analizaremos en primer lugar las 4 obras que componen el Corpus A y a las que consideramos indubitadas, al objeto de disponer de una tabla de valores que reflejen numéricamente el estilo del autor estudiado. De este análisis obtendremos los RdI y el RdA. A continuación analizaremos cada una de las 8 piezas que integran el Corpus B y compararemos los resultados con los datos obtenidos a partir del Corpus A.

Fase 5: Determinación de porcentajes de autoría de las obras del Corpus

B. Por último, el grado de separación de los valores de cada obra del

Corpus B con respecto a las obras del Corpus A, se contrastará con el LAR para determinar cuáles son las que lo superan y por tanto pueden considerarse no atribuibles a Carlos Castaneda.

4.4.2.1. Fase 1: Selección de los marcadores estilísticos

Al objeto de establecer los Rasgos de Identidad estilística del escritor Carlos Castaneda precisamos de unos marcadores adecuados y tras diversas pruebas⁶⁷ decidimos recurrir a la frecuencia de aparición de palabras y de n-gramas de caracteres. Ambos marcadores han sido suficientemente testados en estudios previos, generando resultados que avalan su utilización en el presente trabajo (Blasco y Ruíz, 2009; Grieve, 2007; Keselj et al., 2013; Spassova, 2009).

En cuanto a los marcadores individuales en concreto, estos habitualmente se obtienen extrayéndolos de las obras analizadas. Sin embargo, para nuestro trabajo hemos preferido guiarnos por aquellos que aparecen con más frecuencia en los textos literarios en general. Para ello nos ha sido de gran utilidad el estudio realizado por Peter Norvig, de los laboratorios Google Inc. (Norvig, 2013), a partir de un corpus de más de cinco millones y medio de libros publicados en lengua inglesa (Lin, Michel, Aiden, Orwant, Brockman y Petrov, 2012). Norvig había retomado la idea que en la década de 1960 llevó a Mark Mayzner⁶⁸ a extraer las frecuencias con que aparecen ciertas combinaciones de letras en la lengua inglesa escrita

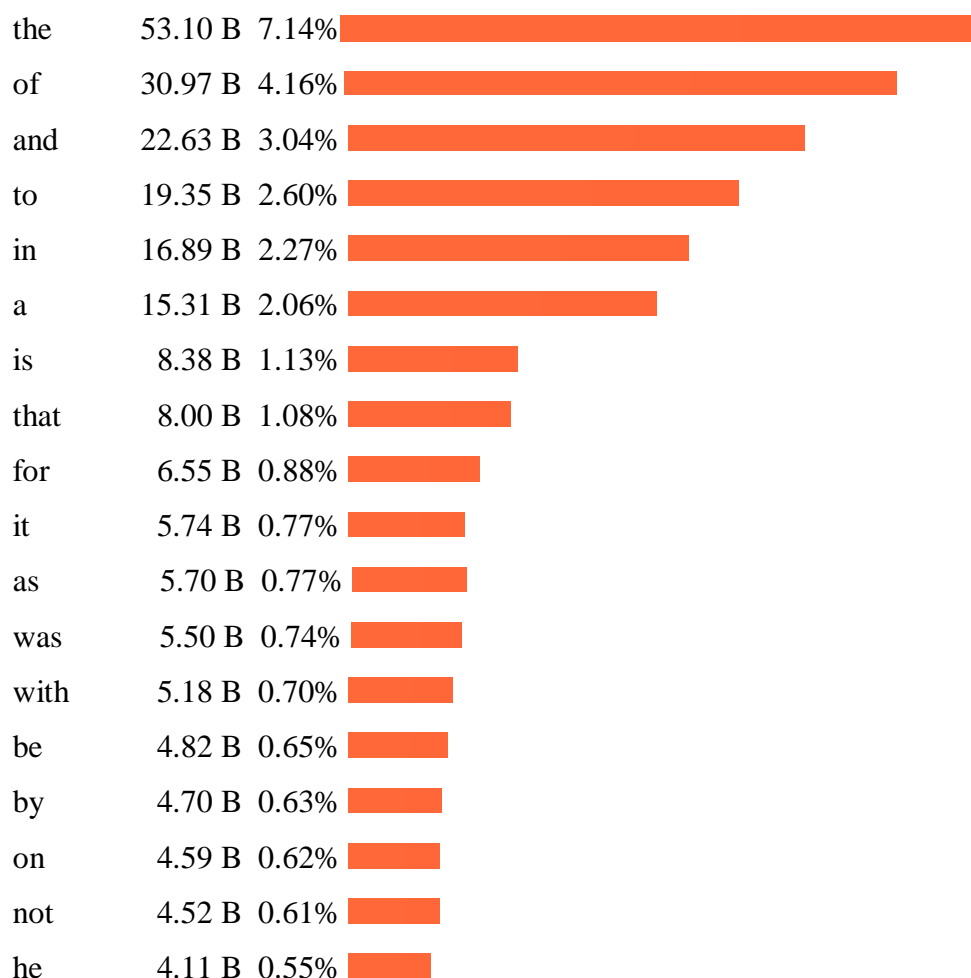
⁶⁷ Previamente habíamos llevado a cabo análisis estilométricos empleando como marcadores estilísticos las categorías gramaticales y las longitudes de palabras. En el primer caso fue preciso encargar una herramienta informática específica a la empresa de tratamiento de lenguaje natural Molino de Ideas, pero ni en este caso, ni en el segundo – las longitudes de palabras – obtuvimos resultados satisfactorios.

⁶⁸ Mayzner había trabajado a partir de un corpus de 20.000 palabras elaborado de fuentes muy diversas (libros, periódicos, revistas, etc.) y había ido extrayendo tanto el número total de palabras de 2, de 3, de 4, de 5, de 6 y de 7 letras, como el número de palabras diferentes que integraban los grupos anteriores. Mayzner había utilizado métodos semi-automáticos, en parte manuales y en parte empleando los rudimentarios dispositivos informáticos basados en tarjetas perforadas (fichas Hollerith) que se usaban en la época. De esta forma Mayzner elaboró tablas de frecuencias de aparición de las palabras de diferentes longitudes.
































(Mayzner, 1965). Tras un elemental proceso de adaptación de los textos (convertirlos a un único tipo de letra y eliminando cualquier otro carácter que no fuese una de las 26 letras del alfabeto latino), Norvig generó una serie de tablas de palabras y de combinaciones de letras, n-gramas, clasificándolas por su longitud y frecuencia de aparición.

En lo relativo a las palabras, este análisis produjo un total de 97.565 elementos distintos, que aparecían 743.842.922.321 veces y de donde extrajo la lista de los 50 vocablos más frecuentes, por orden de mayor a menor y donde se aprecia con claridad el predominio de las palabras funcionales sobre las de contenido léxico:

Gráfico 1:



4. Estudio de las obras atribuidas a Carlos Castaneda

i	3.88 B	0.52%	
this	3.83 B	0.51%	
are	3.70 B	0.50%	
or	3.67 B	0.49%	
his	3.61 B	0.49%	
from	3.47 B	0.47%	
at	3.41 B	0.46%	
which	3.14 B	0.42%	
but	2.79 B	0.38%	
have	2.78 B	0.37%	
an	2.73 B	0.37%	
had	2.62 B	0.35%	
they	2.46 B	0.33%	
you	2.34 B	0.31%	
were	2.27 B	0.31%	
their	2.15 B	0.29%	
one	2.15 B	0.29%	
all	2.06 B	0.28%	
we	2.06 B	0.28%	
can	1.67 B	0.22%	
her	1.63 B	0.22%	
has	1.63 B	0.22%	
there	1.62 B	0.22%	
been	1.62 B	0.22%	
if	1.56 B	0.21%	
more	1.55 B	0.21%	
when	1.52 B	0.20%	
will	1.49 B	0.20%	
would	1.47 B	0.20%	
who	1.46 B	0.20%	
so	1.45 B	0.19%	
no	1.40 B	0.19%	

Los 50 vocablos más frecuentes de la lengua inglesa

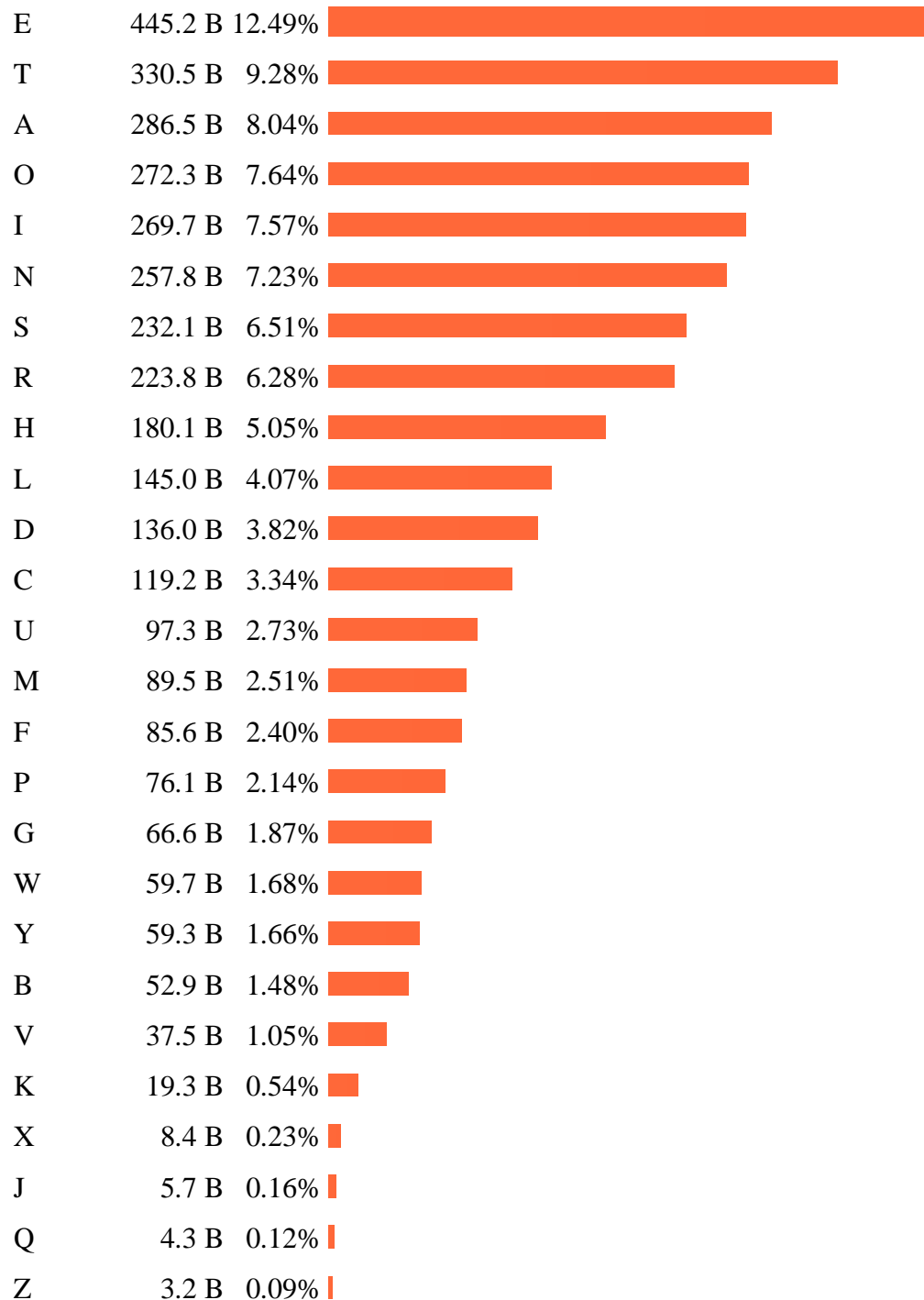
Por nuestra parte, razonamos que estos podrían ser buenos elementos de referencia a la hora de buscar los Rasgos de Identidad estilística del escritor analizado, Castaneda. Dado que vamos a examinar tanto las obras que consideramos indubitadas – Corpus A – como las que nos merecen la consideración de dubitadas – Corpus B – y que pueden, a priori, haber sido escritas por cualquier otro autor, tomaremos como marcadores estilísticos los listados de palabras y n-gramas de aparición más frecuente en un corpus que abarca una enorme población de escritores de ficción en lengua inglesa.

Al hacerlo así, generaremos unos elementos “neutros” que nos permitirán contrastar los textos de estudio con mayor objetividad y, además, nos aseguramos de que encontraremos estos elementos con facilidad.

Por otro lado, dado que nuestro trabajo se apoya en elementos automáticos, pero que implican el uso de técnicas estadísticas sencillas que pudieran generar dudas acerca de su eficacia al ser comparadas con aquellas más sofisticadas que son habituales en trabajos similares, reforzaremos la fiabilidad de los resultados incrementando sustancialmente el número de marcadores estilísticos empleados. Mientras que estudios de atribución de autoría en casos de candidato único toman seis, ocho, doce o, en ciertas ocasiones, cincuenta marcadores (Binongo, 2003; Grant, 2007; Grieve 2007), nosotros emplearemos en un primer momento 276, de entre los que seleccionaremos aquellos de mayor utilidad para nuestros propósitos.

Para ello, y como decíamos antes, nos ha sido de gran utilidad el estudio de Norvig, de donde hemos tomado tanto el listado de las 50 palabras más frecuentes anotado más arriba, como el de los 50 N-gramas de 1 (gráfico 2), de 2, de 3, de 4 y de 5 letras, que se muestran a continuación (tabla 6).

Gráfico 2



Listado de n-gramas de 1 letra⁶⁹, ordenado por frecuencia de aparición

⁶⁹ El estudio de Norvig alteraba, por cierto, la tradicional secuencia de las letras de mayor frecuencia de aparición en lengua inglesa empleada por los linotipistas desde hacía casi cien años: ETAOIN SHRDLU. El nuevo orden de frecuencia de aparición establece una nueva combinación, ligeramente diferente: ETAOIN SRHLDCU.

Tabla 6

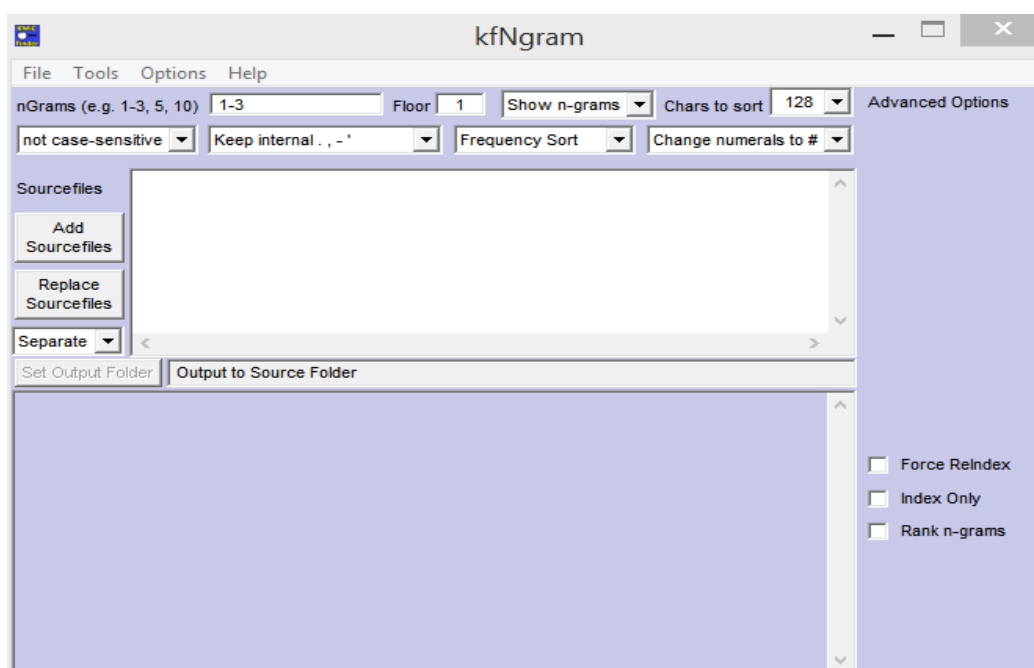
Frecuencia de aparición de N-gramas de 2, de 3, de 4 y de 5 letras				
Nºde orden	2 letras	3 letras	4 letras	5 letras
1	th	the	tion	ation
2	he	and	Atio	tions
3	in	ing	that	which
4	er	ion	ther	ction
5	an	tio	with	other
6	re	ent	ment	their
7	on	ati	ions	there
8	at	for	this	ition
9	en	Her	here	ement
10	nd	ter	from	inter
11	ti	hat	ould	ional
12	es	tha	ting	ratio
13	or	ere	hich	would
14	te	ate	whic	tiona
15	of	his	ctio	these
16	ed	con	ence	state
17	is	res	have	natío
18	it	ver	othe	thing
19	al	all	ight	under
20	ar	ons	sion	ssion
21	st	nce	ever	ectio
22	to	men	ical	catío
23	nt	ith	they	latío
24	ng	ted	inte	about
25	se	ers	ough	count
26	ha	pro	ance	ments
27	as	thi	were	rough
28	ou	wit	tive	ative
29	io	are	over	prese

30	le	ess	ding	feren
31	ve	not	pres	hough
32	co	ive	nter	ution
33	me	was	comp	roduc
34	de	ect	able	resen
35	hi	rea	heir	thoug
36	ri	com	thei	press
37	ro	eve	ally	first
38	ic	per	ated	after
39	ne	int	ring	cause
40	ea	est	ture	where
41	ra	sta	cont	tatio
42	ce	cti	ents	could
43	li	ica	cons	efore
44	ch	ist	rati	contr
45	ll	ear	thin	hould
46	be	ain	part	shoul
47	ma	one	form	tical
48	si	our	ning	gener
49	om	iti	ecti	esent
50	ur	rat	some	great

4.4.2.2. Fase 2: Descripción de las herramientas informáticas

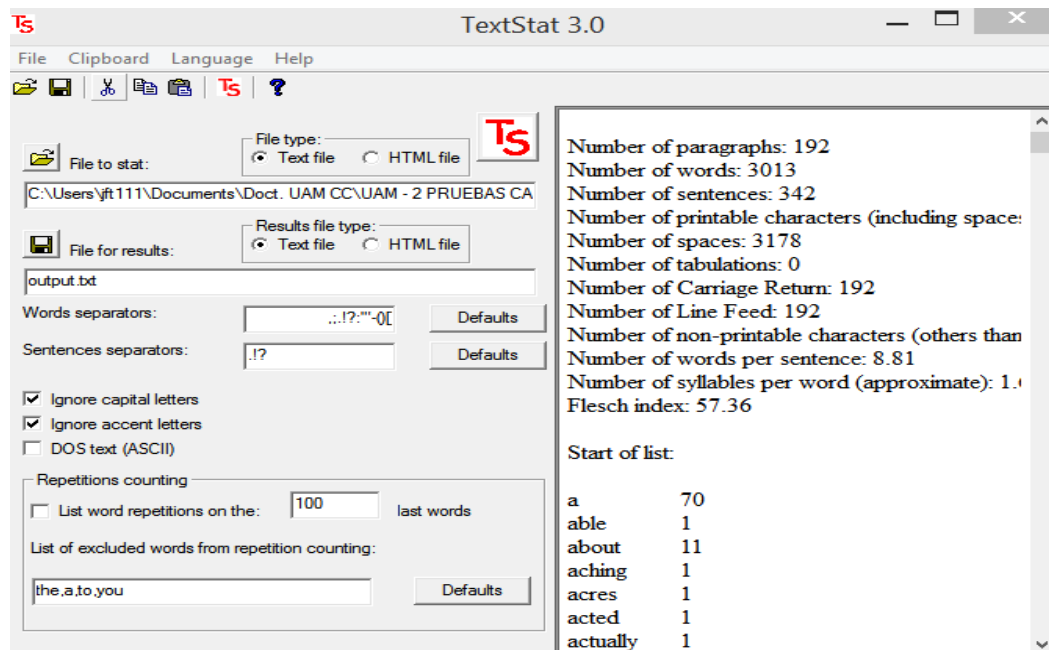
Una vez elegidos los marcadores que se utilizarían, el siguiente paso consistió en determinar las herramientas de procesamiento de texto más adecuadas a nuestros fines. Actualmente están disponibles una variedad de programas que se acercan a las funciones de extracción de datos que necesitaremos para llevar a cabo los análisis contrastivos. De entre ellas merece la pena destacar las siguientes:

a. **kfNgram**. Elaborado por William H. Fletcher (2012), genera listados de n-gramas a partir de textos y archivos HTML. Este programa ofrece una interfaz de uso intuitivo, por lo que el aprendizaje de su manejo resulta relativamente sencillo. Es posible programar la búsqueda por número mínimo de apariciones y por longitud, o número de caracteres de los n-gramas. A pesar de que está preparado para trabajar con el alfabeto latino, es posible configurarlo para que reconozca otros caracteres.

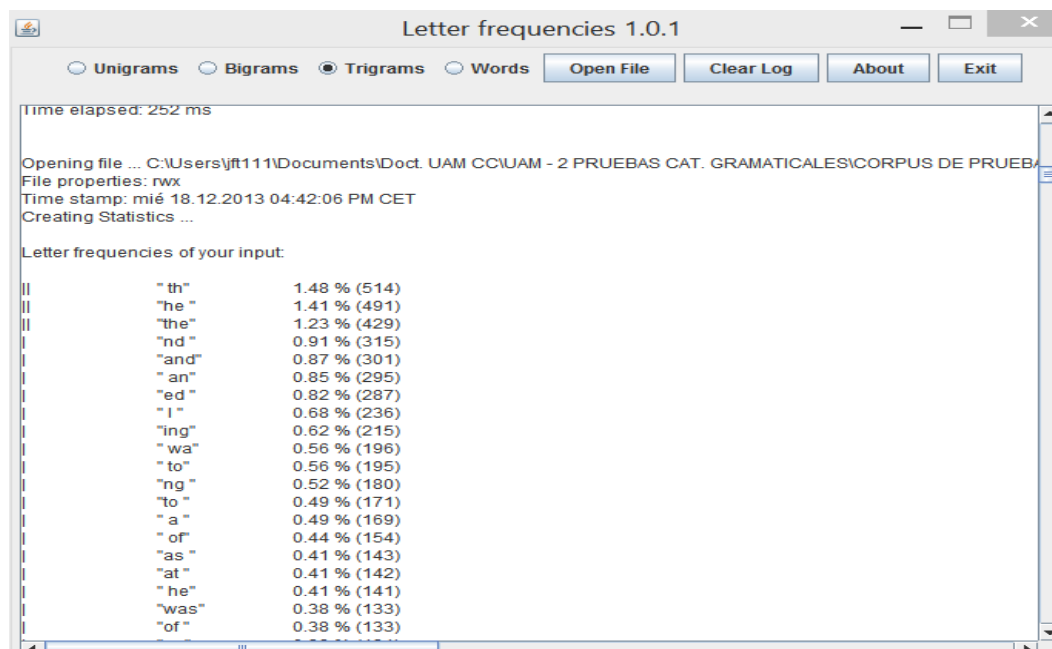


b. **TextStat**. Se trata de un sencillo programa de tratamiento de textos a partir de archivos en texto plano y HTML que genera listas de palabras y concordancias entre distintos archivos. Ha sido desarrollado por el Departamento de Lingüística de la Universidad holandesa de Neon (2005) y es de libre disposición a través de internet. Es, así mismo, un programa multilingüe gracias al uso de Unicode.

4. Estudio de las obras atribuidas a Carlos Castaneda



c. **LetterFrequency**. Desarrollado por Daniel Borkman (2011). De nuevo un programa que permite obtener frecuencias de aparición de palabras, de manejo todavía más sencillo que el anterior, si bien de prestaciones más limitadas.



d. WordSmith 6.0. Sin duda el más completo de todos ellos; su manejo es complejo, pero es también el más eficiente y versátil. Desarrollado por Mike Scott y publicado por Oxford University Press (1996), dispone de una serie de parámetros de búsqueda muy diversos, que permiten acceder a porcentajes de aparición, longitudes medias, *Type-Token Ratios*, etc. Las posibilidades que ofrece hacen que merezca la pena el esfuerzo que supone el aprendizaje y por otro lado, su creador, Mike Scott, está disponible para consultas al respecto. Es este programa el que hemos empleado principalmente en nuestro análisis.

N	Word	Freq.	%	Texts	%	L	S
1	THE	3.490	5,54	1	100,00		
2	I	2.418	3,84	1	100,00		
3	TO	2.031	3,22	1	100,00		
4	AND	1.578	2,51	1	100,00		
5	A	1.476	2,34	1	100,00		
6	OF	1.398	2,22	1	100,00		
7	HE	1.300	2,06	1	100,00		
8	IT	1.169	1,86	1	100,00		
9	WAS	1.106	1,76	1	100,00		
10	YOU	931	1,48	1	100,00		
11	THAT	878	1,39	1	100,00		
12	MY	835	1,33	1	100,00		
13	ME	766	1,22	1	100,00		
14	IN	737	1,17	1	100,00		
15	HAD	652	1,04	1	100,00		

e. Office Word 2013. Por simple que parezca, el sistema de búsqueda de este procesador de texto comercial cumple su función con una precisión y rapidez envidiables. Por desgracia su adaptabilidad a nuestras necesidades es prácticamente nula, lo que ha obligado a

realizar extracciones individuales, elemento a elemento, con la consiguiente inversión de tiempo y esfuerzo.

4.4.2.3. Fase 3: Pruebas con el Corpus de Entrenamiento

Antes de aplicar en los textos objeto de estudio el algoritmo de atribución de autoría que hemos diseñado para casos como el presente, lo probaremos en un corpus constituido por 30 obras de ficción, escritas por un total de 10 autores, a razón de 3 obras por autor. Los textos están tomados de la Biblioteca Gutenberg y pertenecen a escritores de lengua inglesa que desarrollaron su labor creativa dentro del género novelístico y en el mismo período temporal. Es decir, todos ellos surgen del mismo contexto cultural e histórico, por lo que al compararlos entre sí, es de esperar una mayor proximidad estilística de la que existiría si se tomaran textos producidos en épocas muy distantes entre sí, o si se comparasen géneros dispares, como la novela y teatro o drama y poesía.

Por otro lado, del análisis que sigue obtendremos un LAR aplicable a conjuntos de obras que mantengan condiciones similares de homogeneidad, como las que constituirán el objeto de nuestro estudio. Podemos afirmar con suficiente convicción que el conjunto de las obras atribuidas a Carlos Castaneda se ajustan a dichas condiciones, pues han sido producidas en la misma época, pertenecen todas ellas al mismo género (sea este cual sea) e incluso pertenecen, a priori, al mismo autor.

A continuación listamos la relación de autores y obras que componen el Corpus de Entrenamiento:

CORPUS DE ENTRENAMIENTO. AUTORES.

- 1) Autor 1: **F. Anstey**
 - a) Obra 1: *The Talking Horse* (1892)
 - b) Obra 2: *Baboo Jabberjee, B.A.* (1987)
 - c) Obra 3: *The Brass Bottle* (1900)

- 2) Autor 2: **Lucy Maud Montgomery**
 - a) Obra 1: *Anne of Green Gables* (1908)
 - b) Obra 2: *Chronicles of Avonlea* (1912)
 - c) Obra 3: *Rainbow Valley* (1919)

- 3) Autor 3: **Alice Brown**
 - a) Obra 1: *Tiverton Tales* (1899)
 - b) Obra 2: *Country Neighbors* (1910)
 - c) Obra 3: *The Prisoner* (1916)

- 4) Autor 4: **Rudyard Kipling**
 - a) Obra 1: *Soldiers Three* (1888)
 - b) Obra 2: *Under the Deodars* (1888)
 - c) Obra 3: *Kim* (1901)

- 5) Autor 5: **H. G. Wells**
 - a) Obra 1: *The Invisible Man* (1897)
 - b) Obra 2: *The First Men in the Moon* (1909)
 - c) Obra 3: *The History of Mr. Polly* (1910)

- 6) Autor 6: **E. W. Hornung**
 - a) Obra 1: *The Amateur Cracksman* (1899)
 - b) Obra 2: *Raffles. Further Adventures* (1901)
 - c) Obra 3: *Mr. Justice Raffles* (1909)

7) Autor 7: **Jerome K. Jerome**

- a) Obra 1: *Three Men in a Boat* (1889)
- b) Obra 2: *Diary of a Pilgrimage* (1891)
- c) Obra 3: *Three Men on the Bummel* (1900)

8) Autor 8: **Stephen Leacock**

- a) Obra 1: *Arcadian Adventures with the Idle Rich* (1914)
- b) Obra 2: *Further Foolishness* (1916)
- c) Obra 3: *Frenzied Fiction* (1918)

9) Autor 9: **Sax Rohmer**

- a) Obra 1: *The Devil Doctor* (1916)
- b) Obra 2: *Brood of the Witch Queen* (1918)
- c) Obra 3: *Bat Wing* (1921)

10) Autor 10: **P. G. Wodehouse**

- a) Obra 1: *The Man With Two Left Feet* (1917)
- b) Obra 2: *My Man Jeeves* (1919)
- c) Obra 3: *Right Ho, Jeeves* (1934)

El primer paso, por tanto, consiste en obtener los Rasgos de Identidad de cada uno de los autores analizados. Conviene recordar que, a pesar de que partimos de los mismos marcadores en todos los casos (los listados de frecuencias de aparición de palabras y n-gramas proporcionados por el estudio de Norvig), un primer análisis nos indicará cuáles son los de mayor utilidad para cada escritor en concreto. Cada autor mostrará una especial regularidad en ciertos marcadores, que podrán coincidir o no con los de otros escritores.

Veamos, a modo de ejemplo, los valores generados por las obras del Autor 1, F. Anstey, al contrastarlos con las obras del Autor 2, Lucy Maud Montgomery.

En primer lugar obtendremos los valores que arrojan las **50 palabras** de referencia en cada una de las 3 obras del Autor 1, F. Anstey. Las cifras que aparecen en las columnas respectivas (tabla 7) corresponden a porcentajes de aparición de cada marcador concreto dentro del volumen total de palabras en cada obra.

Tabla 7

Comparativa de las tres obras del autor 1, F. Anstey. 50 palabras más frecuentes				
Puesto	Palabra	Obra 1 (%)	Obra 2 (%)	Obra 3 (%)
1	the	5,07	4,12	3,55
2	of	3,66	2,27	1,79
3	and	3,29	2,65	2,83
4	to	2,84	2,87	2,83
5	in	1,61	1,48	1,27
6	a	2,85	2,07	2,20
7	is	0,83	0,55	0,46
8	that	1,75	1,56	1,40
9	for	0,83	0,94	0,88
10	it	0,75	1,61	1,49
11	as	0,78	0,95	0,85
12	was	1,11	1,00	1,51
13	with	1,08	0,80	0,83
14	be	0,50	0,81	0,66
15	by	0,71	0,36	0,34
16	on	0,53	0,53	0,52
17	not	0,67	0,70	0,81
18	he	0,66	1,93	1,85
19	i	2,64	2,58	3,16
20	this	0,35	0,55	0,44
21	are	0,21	0,19	0,28
22	or	0,32	0,23	0,20
23	his	0,57	0,98	0,70

4. Estudio de las obras atribuidas a Carlos Castaneda

24	from	0,38	0,25	0,22
25	at	0,66	0,54	0,62
26	which	0,60	0,40	0,37
27	but	0,55	0,68	0,73
28	have	0,48	0,61	0,57
29	an	0,42	0,32	0,24
30	had	0,44	0,61	0,85
31	they	0,21	0,34	0,51
32	you	0,51	1,93	1,92
33	were	0,26	0,25	0,31
34	their	0,21	0,11	0,16
35	one	0,17	0,26	0,27
36	all	0,30	0,55	0,61
37	we	0,13	0,16	0,33
38	can	0,12	0,33	0,25
39	her	0,48	0,29	0,75
40	has	0,20	0,09	0,11
41	there	0,13	0,33	0,32
42	been	0,13	0,23	0,30
43	if	0,24	0,53	0,38
44	more	0,20	0,25	0,22
45	when	0,20	0,18	0,26
46	will	0,23	0,29	0,28
47	would	0,25	0,43	0,37
48	who	0,28	0,19	0,20
49	so	0,36	0,43	0,54
50	no	0,21	0,41	0,35

A continuación (tabla 8), seleccionaremos aquellos 10 marcadores que menor oscilación presentan de una obra a otra. En la tabla siguiente se muestran los marcadores de la categoría de “Palabras” que constituyen los RdI del Autor 1, F.Anstey.

Tabla 8

Marcadores RdI, categoría “palabra”, de las tres obras de F. Anstey					
Puesto	Palabra	Obra 1 (%)	Obra 2 (%)	Obra 3 (%)	Oscilación(%)
1	to	2,84	2,87	2,83	0,04
2	on	0,53	0,53	0,52	0,01
3	are	0,21	0,19	0,28	0,09
4	were	0,26	0,25	0,31	0,06
5	their	0,21	0,11	0,16	0,10
6	one	0,17	0,26	0,27	0,10
7	more	0,20	0,25	0,22	0,05
8	when	0,20	0,18	0,26	0,08
9	will	0,23	0,29	0,28	0,06
10	who	0,28	0,19	0,20	0,09

A partir de la oscilación que hemos observado en cada uno de los marcadores de esta categoría, calcularemos el RdA (tabla 9)⁷⁰.

Tabla 9

RdA de las obras de F. Anstey		
Puesto	Palabra	Rango de Autoría
1	to	2,79 – 2,91
2	on	0,51 – 0,54
3	are	0,10 – 0,37
4	were	0,20 – 0,37
5	their	0,01 – 0,31
6	one	0,07 – 0,37
7	more	0,15 – 0,30
8	when	0,10 – 0,34
9	will	0,17 – 0,34
10	who	0,10 – 0,37

⁷⁰ A modo de recordatorio, diremos que obtenemos el Rango de Autoría restando la oscilación al valor menor y sumándosela al valor mayor. En el caso del marcador “to”, el RdA tendrá, pues, los valores 2,79 (2,83 – 0,04) y 2,91 (2,87 + 0,04).

Repetiremos la misma operación para cada categoría de marcadores, 1 n-gramas, 2 n-gramas, 3 n-gramas,...etc.

Con los RdA ya establecidos en todas las categorías, el siguiente paso será comprobar cuántas infracciones genera cada obra de los otros nueve autores en cada una de las distintas categorías de marcadores.

Siguiendo nuestro ejemplo, mostramos a continuación (tabla 10), una comparación entre los valores obtenidos por cada una de las 3 obras del Autor 2 con el RdA del Autor 1, para observar cuántas *infracciones* generan, en la categoría de **palabras**.

Tabla 10

Comparación de las obras de F. Anstey y de L.M. Montgomery: infracción de la categoría “palabras”					
MARCADOR	RdA Autor 1	Autor2 Obra1	Autor2 Obra2	Autor2 Obra 3	Infracciones
to	2,79 – 2,91	2,84	2,7	2,58	2
on	0,51 – 0,54	0,53	0,54	0,54	0
are	0,10 – 0,37	0,24	0,16	0,22	0
were	0,20 – 0,37	0,29	0,28	0,36	0
their	0,01 – 0,31	0,1	0,07	0,16	0
one	0,07 – 0,37	0,25	0,23	0,26	0
more	0,15 – 0,30	0,15	0,14	0,13	2
when	0,10 – 0,34	0,42	0,38	0,34	2
will	0,17 – 0,34	0,15	0,15	0,2	2
who	0,10 – 0,37	0,16	0,21	0,18	0
TOTAL		2	4	2	8

Después repetimos la misma operación con cada una de las otras 5 categorías de marcadores: 1 n-grama, 2 n-grama, 3 n-grama,...etc. hasta computar el número de infracciones cometidas por cada obra del Autor 2, que

son aquellos casos en los que o bien superan, o bien no alcanzan los valores del RdA del Autor 1.

La tabla siguiente (tabla 11) muestra las infracciones cometidas por cada una de las Obras del Autor 2 en cada una de las otras 5 categorías de marcadores.

Tabla 11

Infracciones de las obras de L.M. Montgomery en las 5 categorías de marcadores				
MARCADOR	Infracciones Autor2 Obra1	Infracciones Autor2 Obra2	Infracciones Autor2 Obra 3	TOTAL infracciones Autor 2
1 n-gramas	1	2	1	4
2 n-gramas	5	2	4	11
3 n-gramas	2	2	3	7
4 n-gramas	3	2	3	8
5 n-gramas	1	0	1	2
TOTAL	12	8	12	32

Y seguiremos el mismo proceso con cada una de las obras, de cada uno de los otros nueve autores (3 obras por autor) y obtendremos así el valor medio de infracciones que genera cada categoría de marcadores del Autor 1 al compararlo con obras que no han sido escritas por él.

Tabla 12

Categoría	Promedio de infracciones generadas por el Autor 1
1ngramas	3,37
2ngramas	3,6
3ngramas	4,07
4ngramas	4,77
5ngramas	3,85
Palabras	3,37

Es decir, que al comparar el RdA de F. Anstey (Autor1) en la categoría de 1 n-grama, por ejemplo, con una obra de otro autor cualquiera, es de esperar que cometa al menos 3,37 infracciones; en la categoría de 2 n-gramas cometerá 3,6 infracciones como mínimo; en la de 3 n-gramas, obtendremos 4,07 o más infracciones, etc.

A continuación, repetiremos todo el proceso, pero esta vez tomando como referencia el Autor 2, el Autor 3, etc. hasta que cada uno de los diez autores hayan sido comparados con las obras de los otros nueve, en cada una de las 6 categorías de marcadores, lo que da lugar, finalmente, a una tabla de valores (tabla 13) que recoge las infracciones medias cometidas por cada autor en cada categoría.

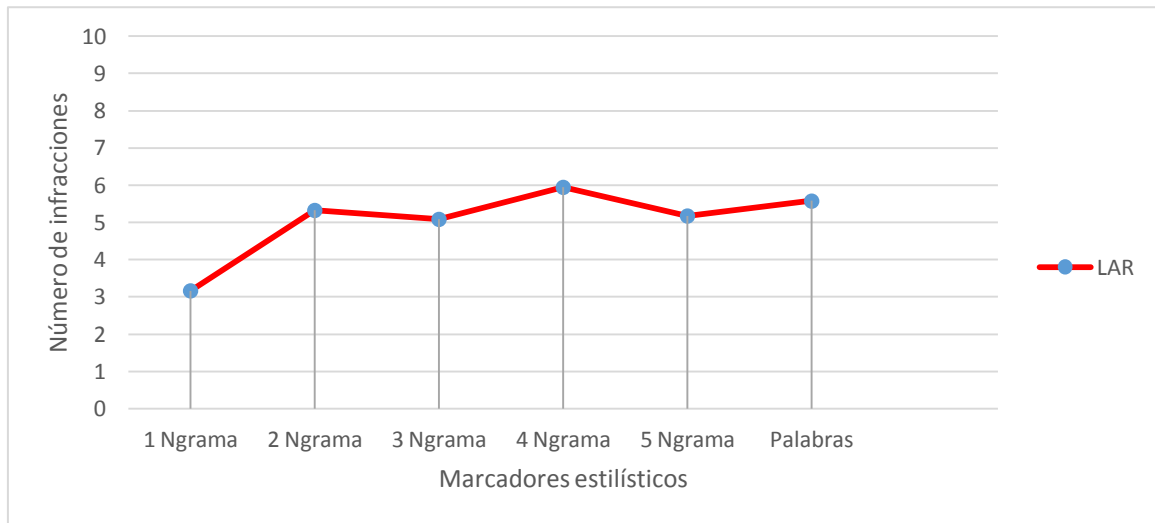
Tabla 13

Media de infracciones de los 10 autores en sus tres obras por cada categoría											
Autores	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Media
1ngramas	3,37	3,29	3,70	2,59	2,50	3,22	3,22	3,07	2,93	3,85	3,17
2ngramas	3,6	4,7	5,5	3,8	6,3	6,9	4,8	4,9	6,1	6,0	5,33
3ngramas	4,07	5,59	5,74	3,96	4,63	5,67	4,40	4,44	6,74	5,70	5,09
4ngramas	4,77	6,18	7	5,07	5,62	7,07	4,85	6,62	5,88	6,48	5,95
5ngramas	3,85	5,74	5,85	5,11	5,88	4,96	3,70	5,29	6,81	4,37	5,18
Palabras	3,37	7,88	4,85	5,29	6,03	5,77	5,37	3,07	6,92	5,22	5,58

Los valores medios que aparecen en la última columna de la derecha son, por tanto, de aplicación general y representan la cantidad media de infracciones que genera una obra escrita por un autor cualquiera cuando se la compara con la supuesta obra de otro escritor diferente.

Son estos valores los que nos permitirán elaborar la línea del LAR. Sabremos así, cuál es el valor medio en cada tipo de marcador, sobrepasado el cual podremos afirmar con bastante certeza que se trata de una obra escrita por otro autor.

Gráfico 3



Línea LAR de marcadores estilísticos

Este será, por tanto, el proceso que seguiremos en el análisis de cada una de las obras dubitadas objeto de nuestro estudio: Primero obtendremos los *RdI* del escritor Carlos Castaneda extrayéndolos de las 4 obras indubitadas – el Corpus A – para a partir de ellos elaborar los *RdA* de cada marcador. A continuación obtendremos los valores de esos mismos Rasgos de Identidad de las 8 obras dubitadas – el Corpus B – y los compararemos anotando el número de *infracciones* que generen. Por último, contrastaremos el número de infracciones cometidas con los valores del LAR que hemos establecido más arriba. A medida que los valores generados por una obra dubitada se acerquen a los del LAR, mayor será la probabilidad de que la obra no haya sido escrita por Carlos Castaneda; si los alcanzan o los sobrepasan, podremos asegurar que dicha obra ha sido producida por otro autor.

4.5. Fase 4: Análisis de las obras indubitadas de Carlos Castaneda. Corpus A

A continuación llevaremos a cabo el análisis de las obras que consideramos indubitadas, a partir del cual esperamos determinar si existen elementos objetivos que nos permitan afirmar que alguna de las obras atribuidas a Carlos Castaneda incluidas en el Corpus B ha sido escrita por un autor diferente.

El primer paso consiste, por tanto, en obtener los RdI a partir de las cuatro obras que componen el Corpus A (tabla 14) y que consideramos indubitadas.

Tabla 14

Corpus A de las obras de Carlos Castaneda
<i>The Teachings of Don Juan: A Yaqui Way of Knowledge</i> (1968)
<i>A Separate Reality: Further Conversations with Don Juan</i> (1971)
<i>Journey to Ixtlan</i> (1972)
<i>Tales of Power</i> (1974)

Para ello, igual que hemos hecho anteriormente, tomaremos aquellos marcadores estilísticos que experimentan una menor variación, es decir, los elementos que se muestran más estables a lo largo de las obras indubitadas (tabla 15).

A partir de estos marcadores y de la oscilación observada, determinaremos el RdA, que es la oscilación máxima que podemos esperar en cualquier texto que haya sido escrito por Castaneda (tablas 16-20).

Para el cálculo del RdA hemos utilizado tres decimales en aquellas ocasiones en que la diferencia entre el valor máximo y el valor mínimo eran muy cercanos. Sin embargo, para el contraste con las obras analizadas emplearemos siempre dos decimales, al objeto de simplificar las operaciones y porque los resultados finales no varían (tabla 21).

Tabla 15

Rasgos de identidad (RdI) corpus A					
1 N-grama	2 N-gramas	3 N-gramas	4 N-gramas	5 N-gramas	Palabras
C	er	and	ions	which	and
S	nd	ion	hich	ction	with
U	of	con	ctio	ition	be
F	is	ons	ight	ratio	by
J	as	nce	sion	state	this
	le	ive	inte	latio	are
	co	was	nter	ments	which
	ri	ica	heir	ution	they
	ic	cti	cons	tatio	you
	ur	rat	rati	gener	has

Tabla 16

Rango de Autoría (RdA) Corpus A. (1 n-grama)				
1 N-grama	Valor mínimo	Valor máximo	Oscilación	RdA
c	2,07	2,21	0,14	1,93 – 2,35
s	5,33	5,53	0,20	5,13 – 5,73
u	2,90	3,04	0,14	2,76 – 3,18
f	1,90	2,07	0,17	1,73 – 2,24
j	0,26	0,30	0,04	0,22 – 0,34

Tabla 17

Rango de Autoría (RdA) Corpus A. (2 n-gramas)				
2 N-grama	Valor mínimo	Valor máximo	Oscilación	RdA
er	1,35	1,39	0,04	1,31 – 1,43
nd	0,99	1,06	0,07	0,92 – 1,13
of	0,57	0,66	0,09	0,48 – 0,75
is	0,62	0,71	0,09	0,53 – 0,80
as	0,82	0,85	0,03	0,79 – 0,88
le	0,56	0,62	0,06	0,50 – 0,68
co	0,35	0,39	0,03	0,32 – 0,42
ri	0,35	0,37	0,02	0,33 – 0,39
ic	0,22	0,26	0,04	0,18 – 0,30
ur	0,32	0,36	0,04	0,28 – 0,40

Tabla 18

Rango de Autoría (RdA) Corpus A. (3 n-gramas)				
3 N-grama	Valor mínimo	Valor máximo	Oscilación	RdA
and	0,70	0,74	0,04	0,66 – 0,78
ion	0,20	0,27	0,07	0,13 – 0,34
con	0,106	1,109	0,003	0,103 – 0,112
ons	0,06	0,07	0,01	0,05 – 0,08
nce	0,12	0,13	0,01	0,11 – 0,14
ive	0,07	0,09	0,02	0,05 – 0,11
was	0,42	0,45	0,03	0,39 – 0,48
ica	0,03	0,04	0,01	0,02 – 0,05
cti	0,04	0,05	0,01	0,03 – 0,06
rat	0,04	0,05	0,01	0,03 – 0,06

Tabla 19

Rango de Autoría (RdA) Corpus A. (4 n-gramas)				
4 N-grama	Valor mínimo	Valor máximo	Oscilación	RdA
ions	0,03	0,04	0,01	0,02 – 0,05
hich	0,03	0,04	0,01	0,02 – 0,05
ctio	0,033	0,037	0,004	0,029 – 0,041
ight	0,10	0,12	0,02	0,08 – 0,14
sion	0,033	0,036	0,003	0,030 – 0,039
inte	0,03	0,04	0,01	0,02 – 0,05
nter	0,02	0,04	0,02	0,00 – 0,06
heir	0,01	0,03	0,02	0,00 – 0,05
cons	0,012	0,017	0,005	0,007 – 0,022
rati	0,01	0,02	0,01	0,00 – 0,03

Tabla 20

Rango de Autoría (RdA) Corpus A. (5 n-gramas)				
5 N-grama	Valor mínimo	Valor máximo	Oscilación	RdA
which	0,03	0,04	0,01	0,02 – 0,05
ction	0,033	0,036	0,003	0,030 – 0,039
ition	0,01	0,02	0,01	0,00 – 0,03
ratio	0,008	0,014	0,006	0,002 – 0,020
state	0,015	0,017	0,002	0,013 – 0,019
latio	0,003	0,009	0,006	0,000 – 0,015
ments	0,00	0,01	0,01	0,00 – 0,02
ution	0,001	0,003	0,002	0,000 – 0,005
tatio	0,006	0,008	0,002	0,004 – 0,010
gener	0,001	0,002	0,001	0,000 – 0,003

Tabla 21

Rango de Autoría (RdA) corpus A (palabras)				
Palabras	Valor mínimo	Valor máximo	Oscilación	RdA
and	2,52	2,70	0,18	2,34 – 2,88
with	0,63	0,70	0,07	0,56 – 0,77
be	0,41	0,46	0,05	0,36 – 0,51
by	0,28	0,37	0,09	0,19 – 0,46
this	0,13	0,23	0,10	0,03 – 0,33
are	0,21	0,33	0,12	0,09 – 0,45
which	0,13	0,17	0,05	0,08 – 0,22
they	0,28	0,35	0,07	0,21 – 0,42
you	1,29	1,58	0,29	1,00 – 1,87
has	0,09	0,14	0,05	0,04 – 0,19

Estos son, por tanto, los valores obtenidos tras el análisis de las cuatro obras que integran el Corpus A y que servirán para realizar el análisis contrastivo de las ocho que integran el Corpus B y que se detallan a continuación.

4.6. Fase 4: Análisis de las obras dubitadas de Carlos Castaneda. Corpus B

En la tabla 22 se proporciona la relación de obras que componen el Corpus B, cuya autoría por Carlos Castaneda está en entredicho:

Tabla 22

Corpus B
<i>The Second Ring of Power</i> (1977)
<i>The Eagle's Gift</i> (1981)
<i>The Fire from Within</i> (1984)
<i>The Power of Silence: Further Lessons of Don Juan</i> (1987)
<i>The Art of Dreaming</i> (1993)
<i>Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico</i> (1998)
<i>The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe</i> (1998)
<i>The Active Side of Infinity</i> (1998)

4.6.1. Análisis de la Obra 1 (Corpus B)

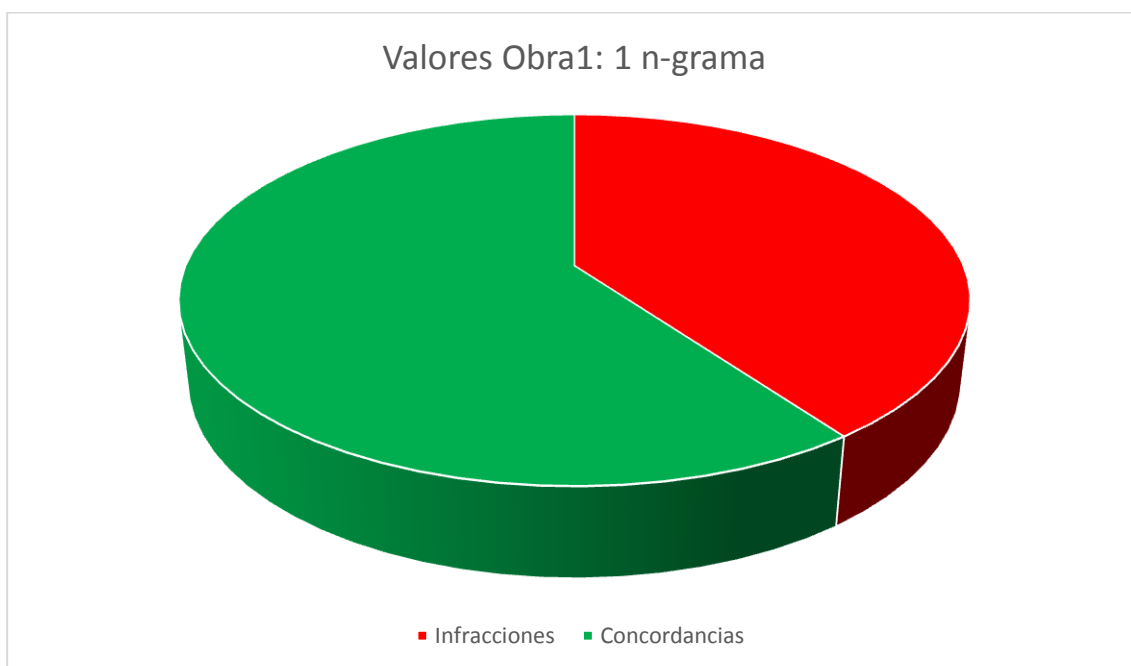
Título de la obra: *The Second Ring of Power*

Fecha de publicación: 1977

Tabla 23

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
“c”	1,93 – 2,35	1,86	X	
“s”	5,13 – 5,73	5,53		X
“u”	2,76 – 3,18	2,79		X
“f”	1,73 – 2,24	1,88		X
“j”	0,22 – 0,34	1,18	X	
TOTAL			2	3

Gráfico 4

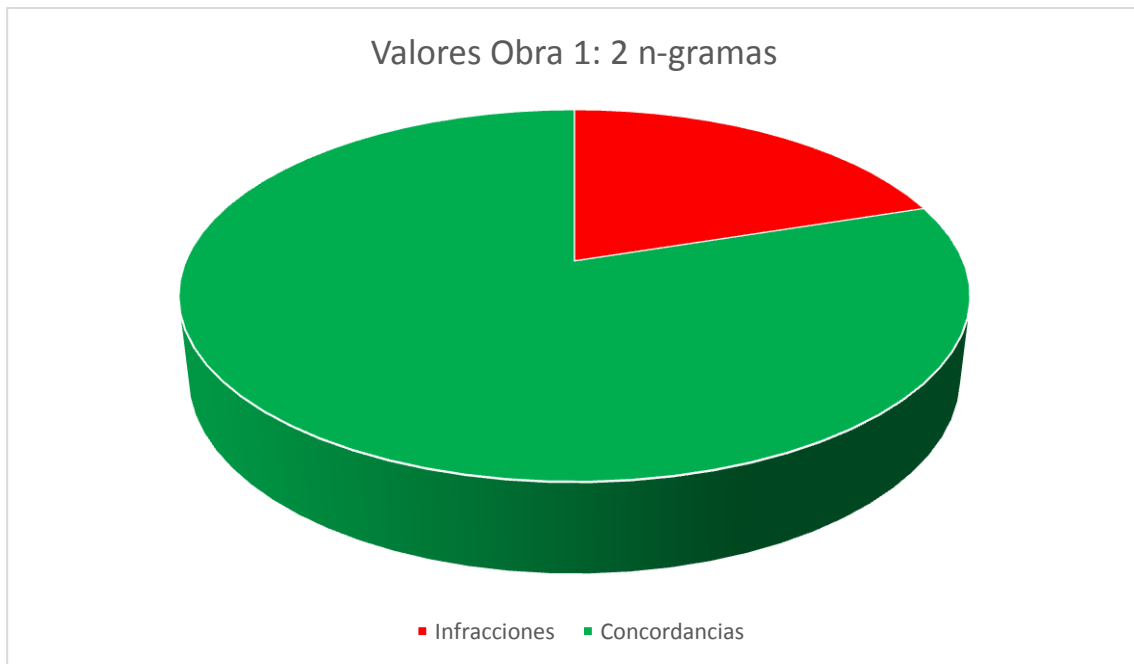


Infracciones y concordancias de 1 n-grama en la obra *The Second Ring of Power*

Tabla 24

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
er	1,31 – 1,43	1,57	X	
nd	0,92 – 1,13	1,04		X
of	0,48 – 0,75	0,54		X
is	0,53 – 0,80	0,41	X	
as	0,79 – 0,88	0,83		X
le	0,50 – 0,68	0,58		X
co	0,32 – 0,42	0,34		X
ri	0,33 – 0,39	0,38		X
ic	0,18 – 0,30	0,28		X
ur	0,28 – 0,40	0,30		X
TOTAL			2	8

Gráfico 5

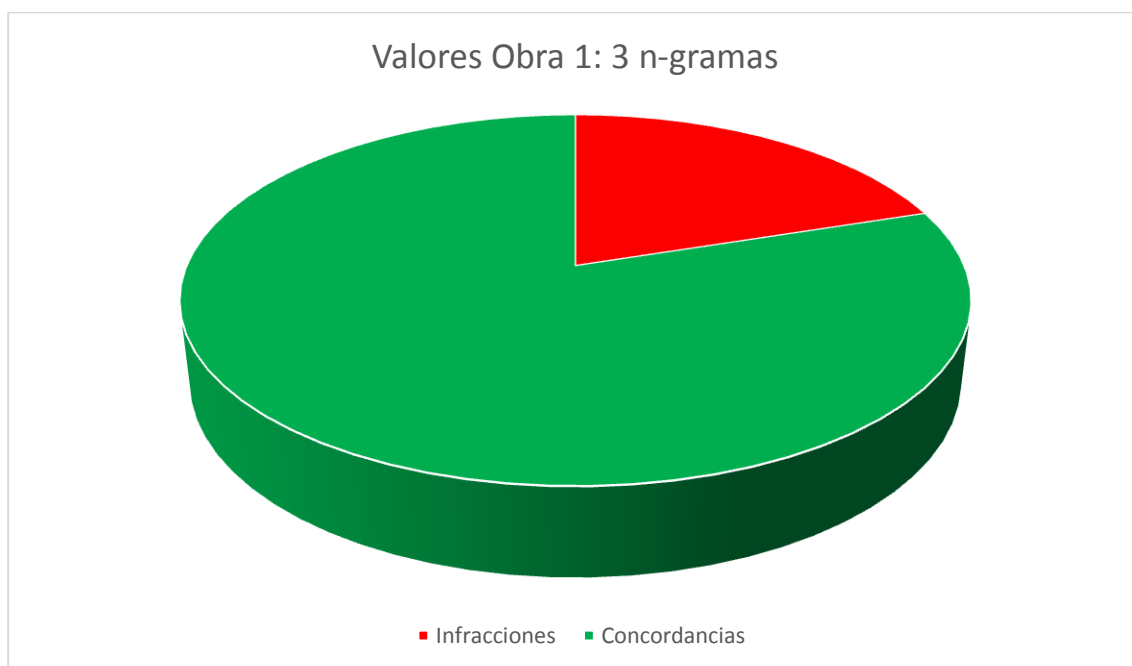


Infracciones y concordancias de 2 n-gramas en la obra *The Second Ring of Power*

Tabla 25

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
and	0,66 – 0,78	0,72		X
ion	0,13 – 0,34	0,20		X
con	0,10 – 0,11	0,08	X	
ons	0,05 – 0,08	0,04	X	
nce	0,11 – 0,14	0,11		X
ive	0,05 – 0,11	0,08		X
was	0,39 – 0,48	0,45		X
ica	0,02 – 0,05	0,02		X
cti	0,03 – 0,06	0,04		X
rat	0,03 – 0,06	0,04		X
TOTAL			2	8

Gráfico 6

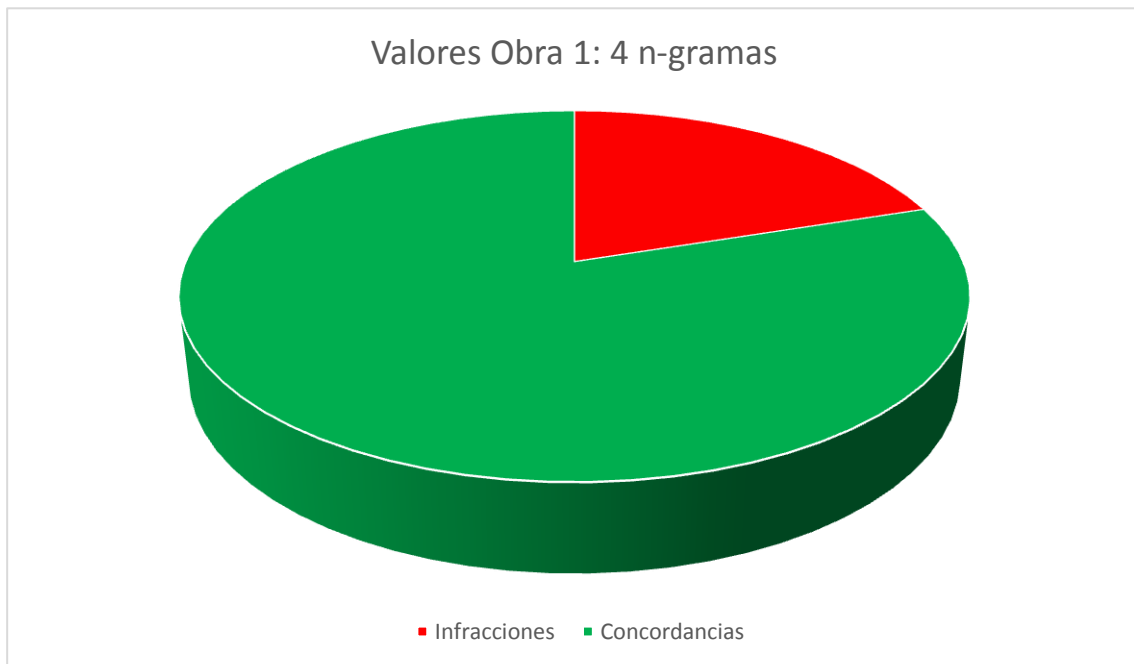


Infracciones y concordancias de 3 n-gramas en la obra *The Second Ring of Power*

Tabla 26

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
ions	0,02 – 0,05	0,03		X
hich	0,02 – 0,05	0,02		X
ctio	0,02 – 0,04	0,03		X
ight	0,08 – 0,14	0,11		X
sion	0,03 – 0,03	0,02	X	
inte	0,02 – 0,05	0,03		X
nter	0,00 – 0,06	0,03		X
heir	0,00 – 0,05	0,06	X	
cons	0,00 – 0,02	0,00		X
рати	0,00 – 0,03	0,01		X
TOTAL			2	8

Gráfico 7

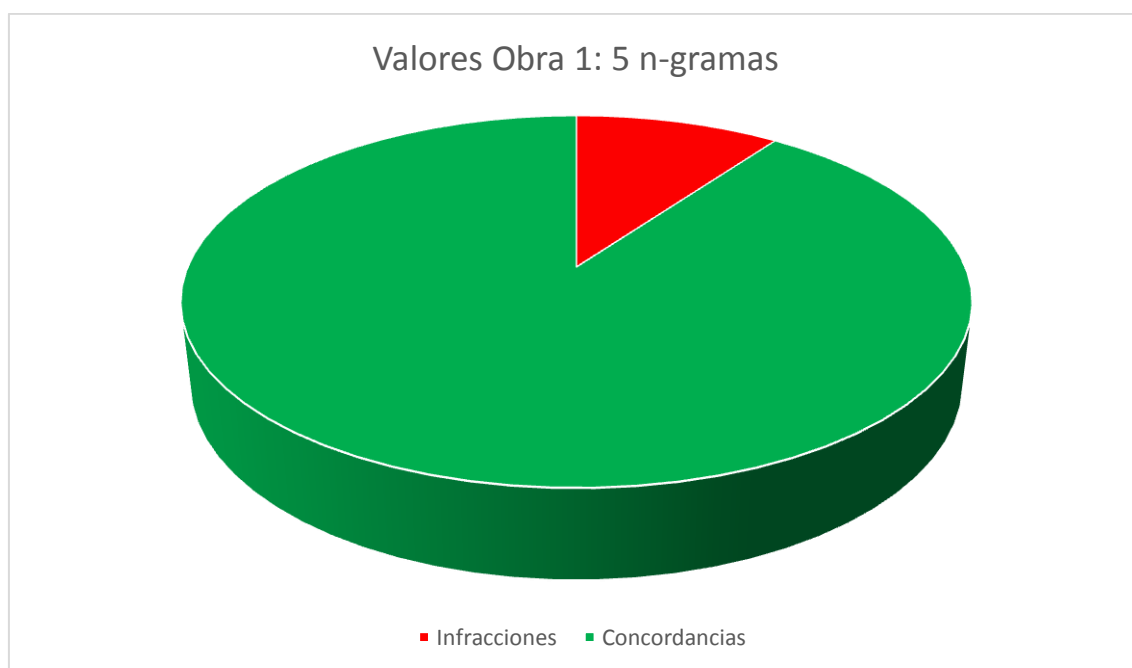


Infracciones y concordancias de 4 n-gramas en la obra *The Second Ring of Power*

Tabla 27

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
which	0,02 – 0,05	0,02		X
ction	0,03 – 0,03	0,03		X
ition	0,00 – 0,03	0,00		X
ratio	0,00 – 0,02	0,01		X
state	0,01 – 0,01	0,00	X	
latio	0,00 – 0,01	0,00		X
ments	0,00 – 0,02	0,00		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,00		X
gener	0,00 – 0,00	0,00		X
TOTAL			1	9

Gráfico 8



Infracciones y concordancias de 5 n-gramas en la obra *The Second Ring of Power*

Tabla 28

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O1	Infracciones	Concordancias
and	2,34 – 2,88	2,56		X
with	0,56 – 0,77	0,72		X
be	0,36 – 0,51	0,37		X
by	0,19 – 0,46	0,26		X
this	0,03 – 0,33	0,09		X
are	0,09 – 0,45	0,03	X	
which	0,08 – 0,22	0,11		X
they	0,21 – 0,42	0,05	X	
you	1,00 – 1,87	1,09		X
has	0,04 – 0,19	1,11	X	
TOTAL			3	7

Gráfico 9



Infracciones y concordancias de palabras en la obra *The Second Ring of Power*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 1 del Corpus B.

Gráfico 10

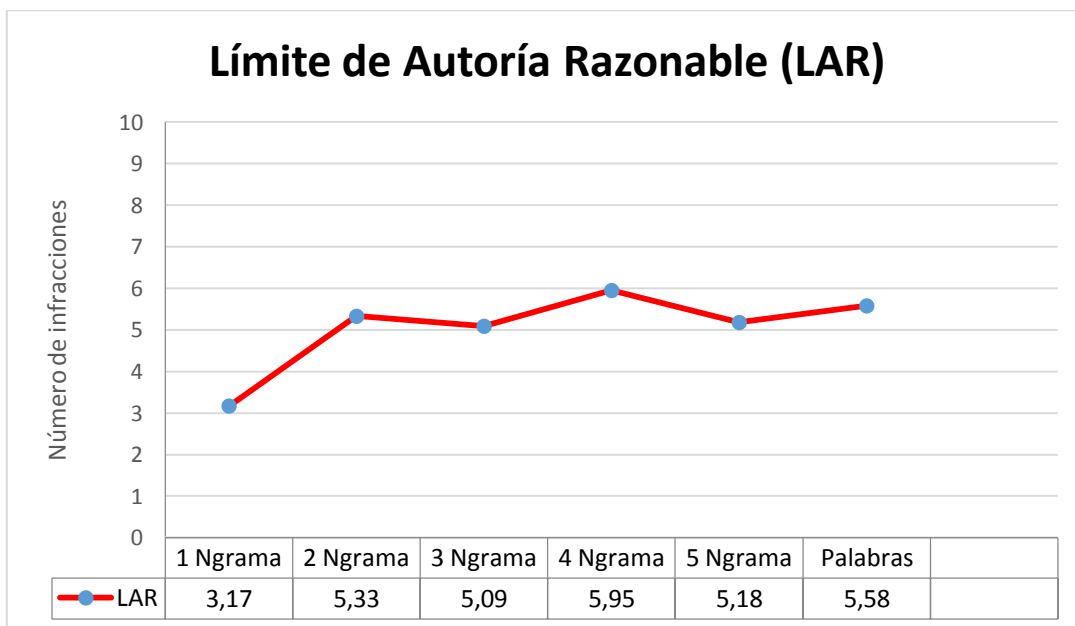


Gráfico 11

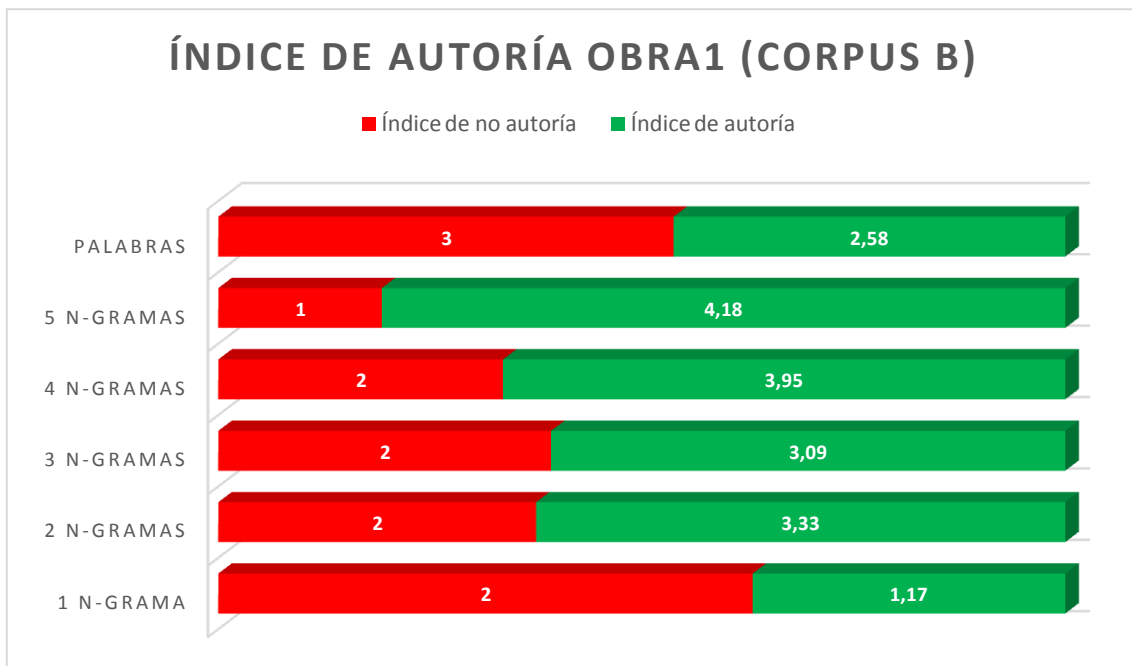


Gráfico 12

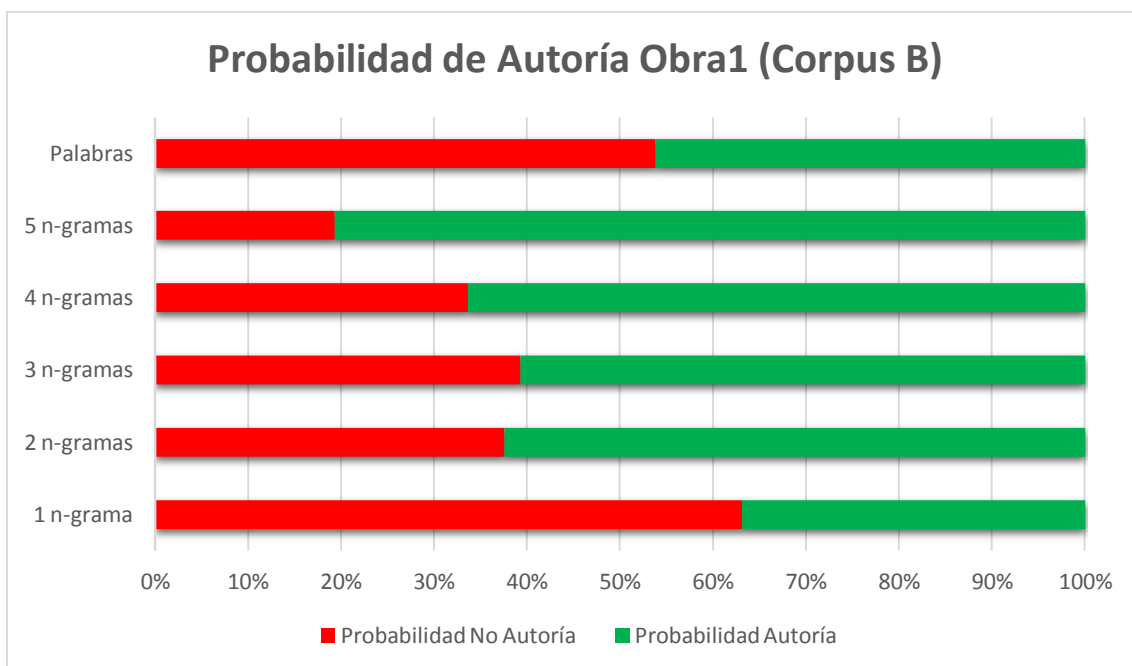
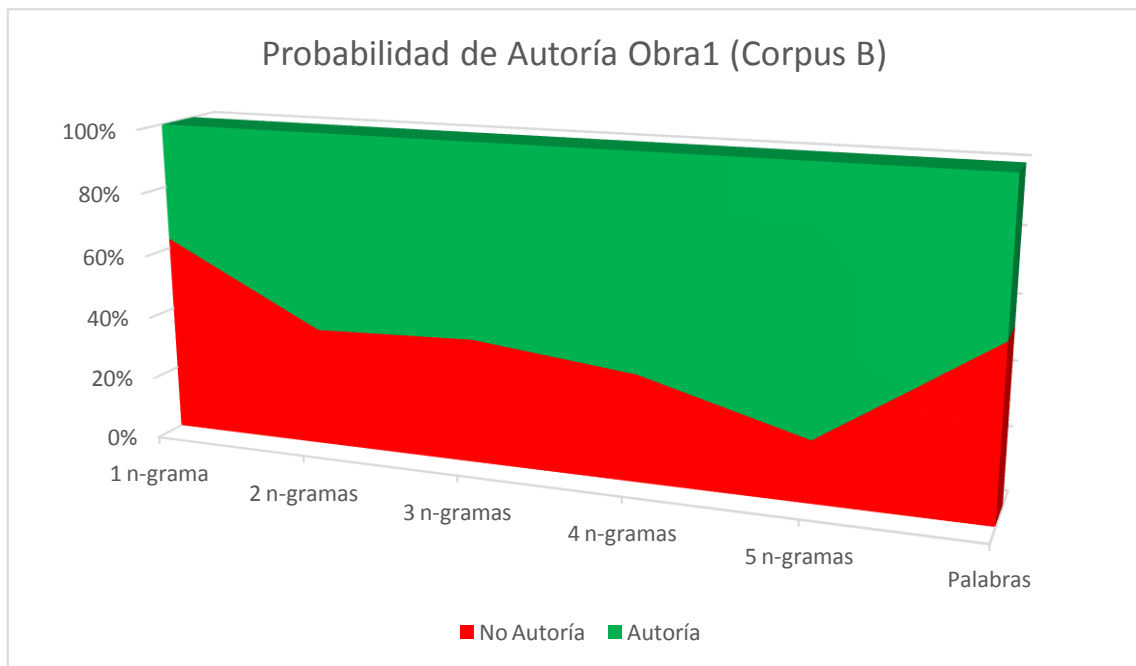


Gráfico 13



4.6.2. Análisis de la Obra 2 (Corpus B)

Título de la obra: *The Eagle's Gift*

Fecha de publicación: 1981

Tabla 29

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
“c”	1,93 – 2,35	2,13		X
“s”	5,13 – 5,73	5,62		X
“u”	2,76 – 3,18	2,71	X	
“f”	1,73 – 2,24	2,06		X
“j”	0,22 – 0,34	0,23		X
TOTAL			1	4

Gráfico 14

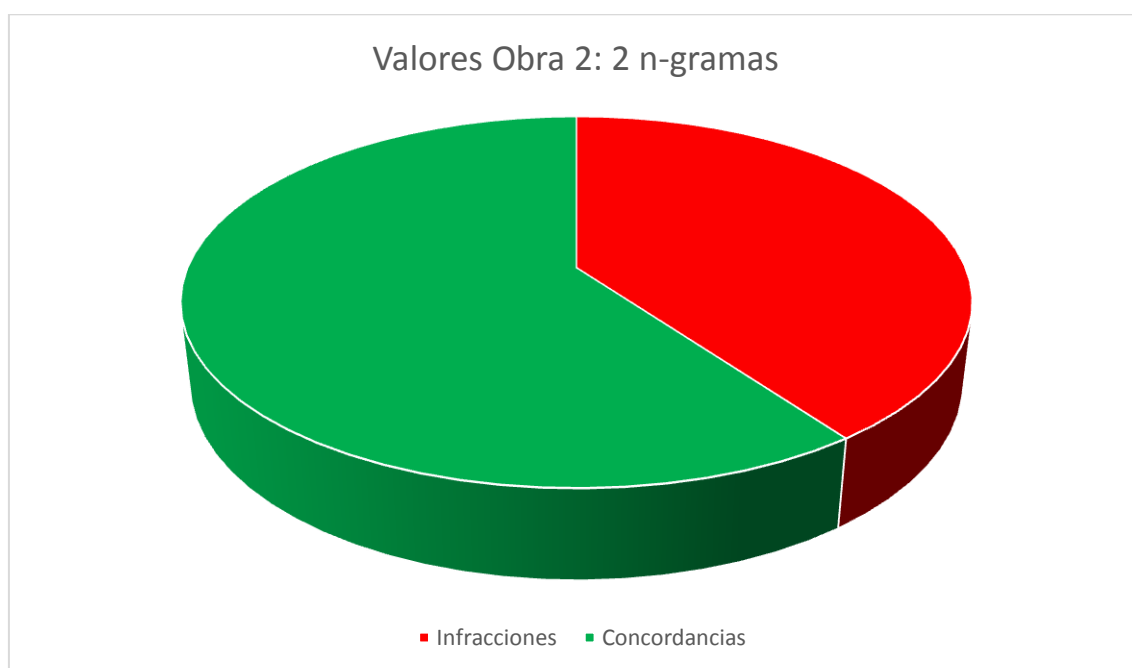


Infracciones y concordancias de 1 n-grama en la obra *The Eagle's Gift*

Tabla 30

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
er	1,31 – 1,43	1,63	X	
nd	0,92 – 1,13	0,92		X
of	0,48 – 0,75	0,66		X
is	0,53 – 0,80	0,44	X	
as	0,79 – 0,88	0,90	X	
le	0,50 – 0,68	0,63		X
co	0,32 – 0,42	0,40		X
ri	0,33 – 0,39	0,41	X	
ic	0,18 – 0,30	0,25		X
ur	0,28 – 0,40	0,34		X
TOTAL			4	6

Gráfico 15

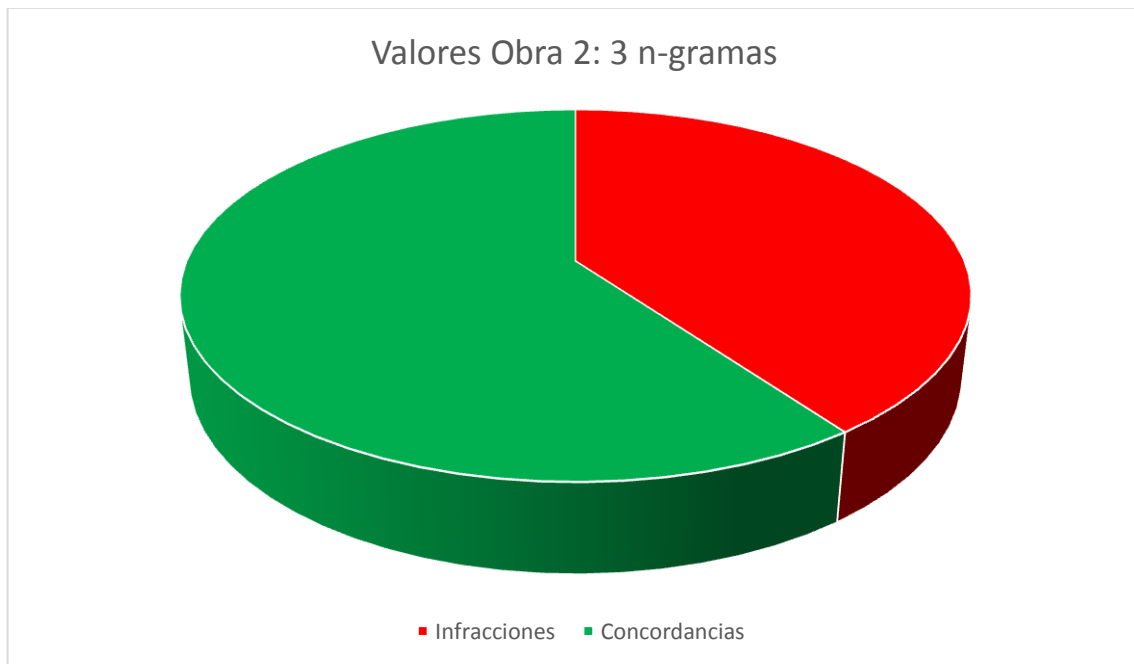


Infracciones y concordancias de 2 n-gramas en la obra *The Eagle's Gift*

Tabla 31

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
and	0,66 – 0,78	0,58	X	
ion	0,13 – 0,34	0,29		X
con	0,10 – 0,11	0,15	X	
ons	0,05 – 0,08	0,08		X
nce	0,11 – 0,14	0,14		X
ive	0,05 – 0,11	0,08		X
was	0,39 – 0,48	0,40		X
ica	0,02 – 0,05	0,07	X	
cti	0,03 – 0,06	0,07	X	
rat	0,03 – 0,06	0,06		X
TOTAL			4	6

Gráfico 16

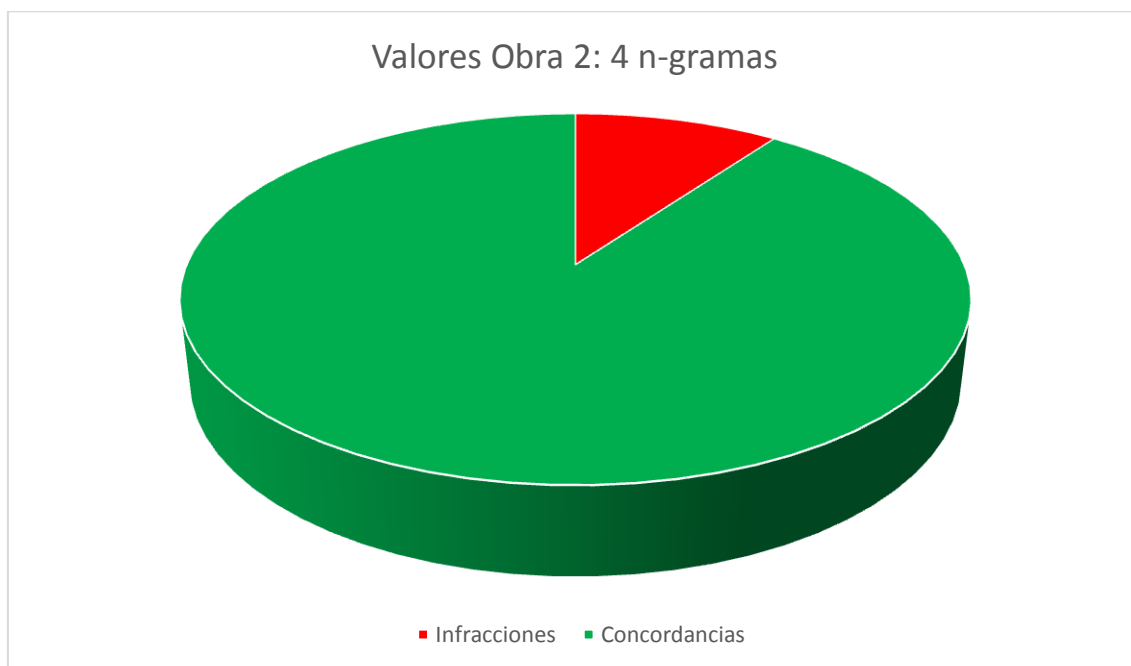


Infracciones y concordancias de 3 n-gramas en la obra *The Eagle's Gift*

Tabla 32

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
ions	0,02 – 0,05	0,05		X
hich	0,02 – 0,05	0,03		X
ctio	0,02 – 0,04	0,04		X
ight	0,08 – 0,14	0,10		X
sion	0,030 – 0,039	0,03		X
inte	0,02 – 0,05	0,05		X
nter	0,00 – 0,06	0,04		X
heir	0,00 – 0,05	0,07	X	
cons	0,007 – 0,022	0,02		X
rati	0,00 – 0,03	0,02		X
TOTAL			1	9

Gráfico 17

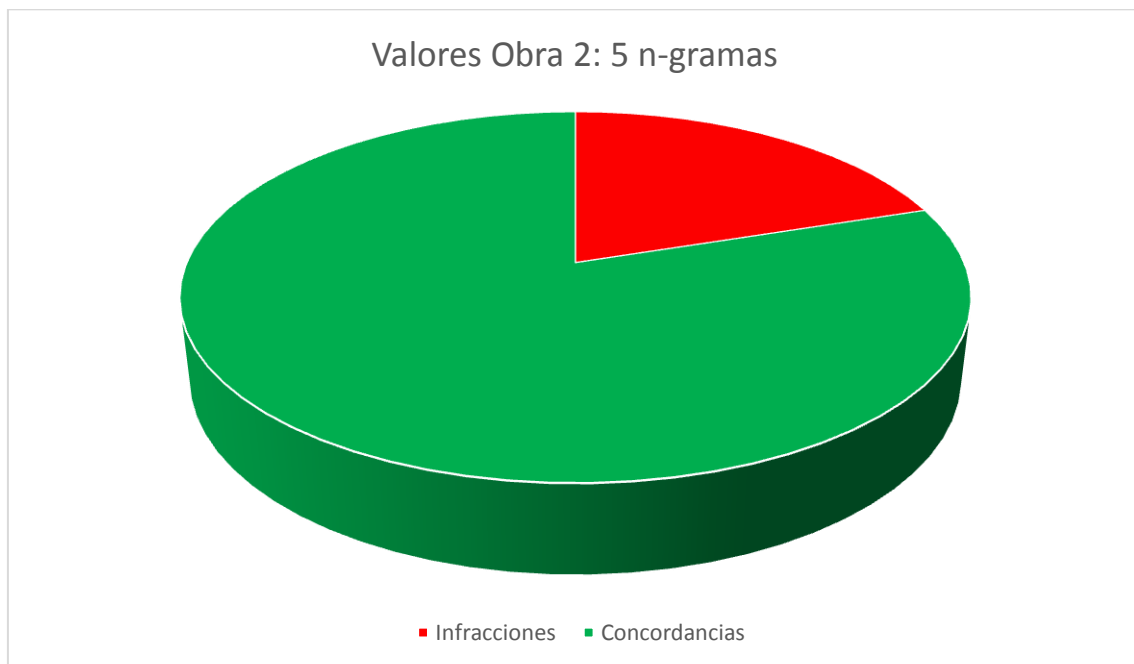


Infracciones y concordancias de 4 n-gramas en la obra *The Eagle's Gift*

Tabla 33

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
which	0,02 – 0,05	0,03		X
ction	0,03 – 0,03	0,04	X	
ition	0,00 – 0,03	0,01		X
ratio	0,00 – 0,02	0,01		X
state	0,01 – 0,01	0,03	X	
latio	0,00 – 0,01	0,00		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,01		X
gener	0,00 – 0,00	0,00		X
TOTAL			2	8

Gráfico 18



Infracciones y concordancias de 5 n-gramas en la obra *The Eagle's Gift*

Tabla 34

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O2	Infracciones	Concordancias
and	2,34 – 2,88	2,26	X	
with	0,56 – 0,77	0,70		X
be	0,36 – 0,51	0,42		X
by	0,19 – 0,46	0,34		X
this	0,03 – 0,33	0,16		X
are	0,09 – 0,45	0,20		X
which	0,08 – 0,22	0,17		X
they	0,21 – 0,42	0,56	X	
you	1,00 – 1,87	0,43	X	
has	0,04 – 0,19	0,08		X
TOTAL			3	7

Gráfico 19



Infracciones y concordancias de palabras en la obra *The Eagle's Gift*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 2 del Corpus B.

Gráfico 20

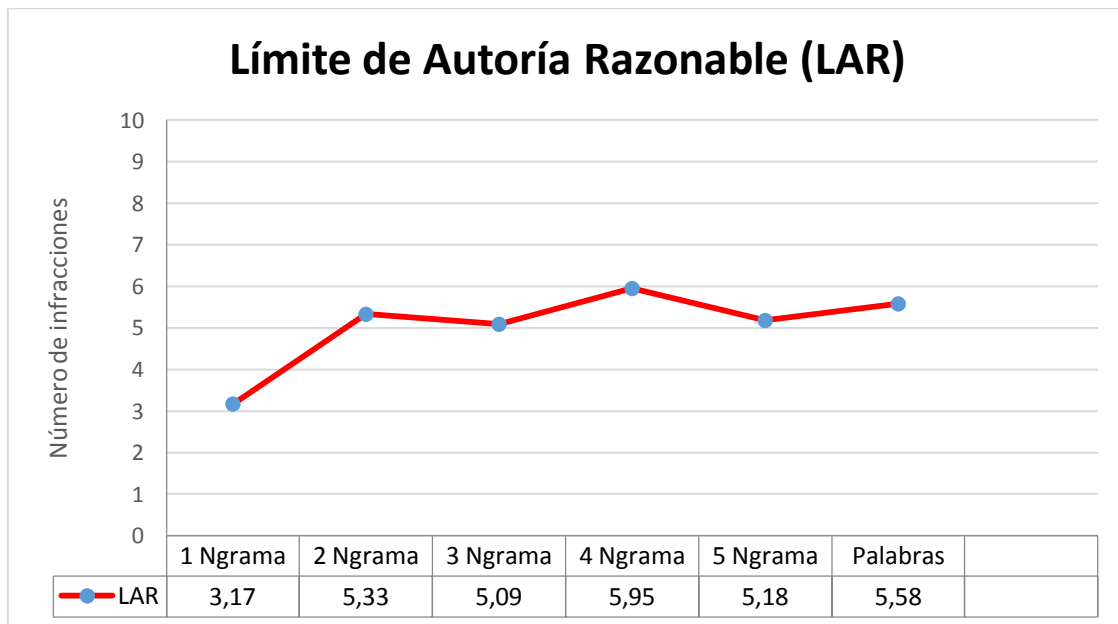


Gráfico 21

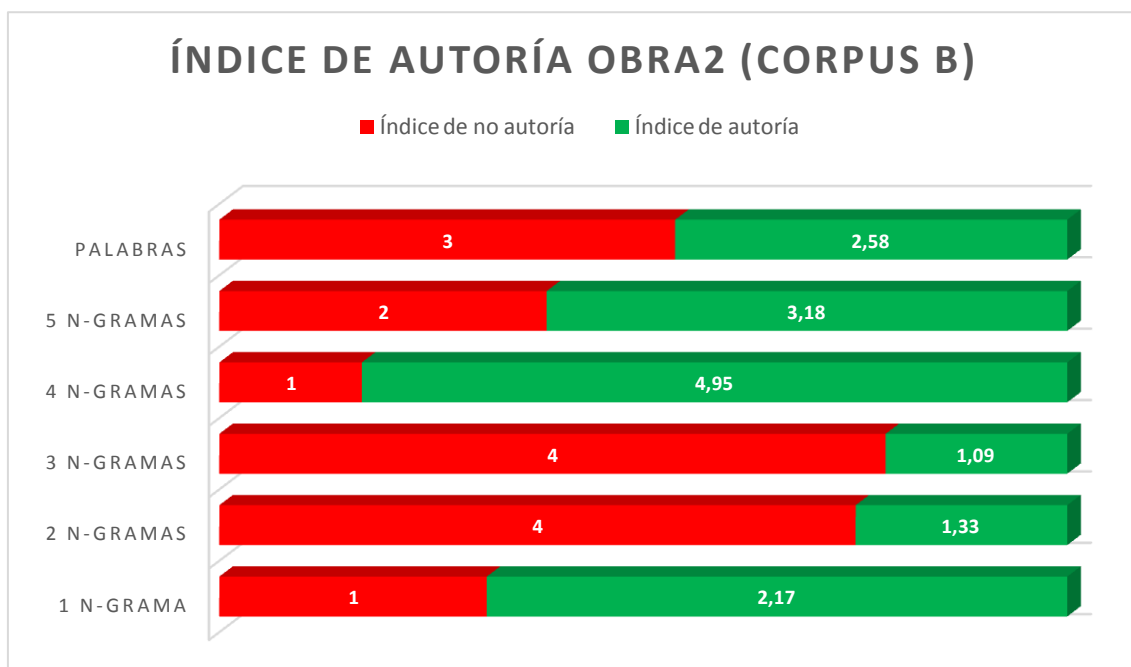


Gráfico 22

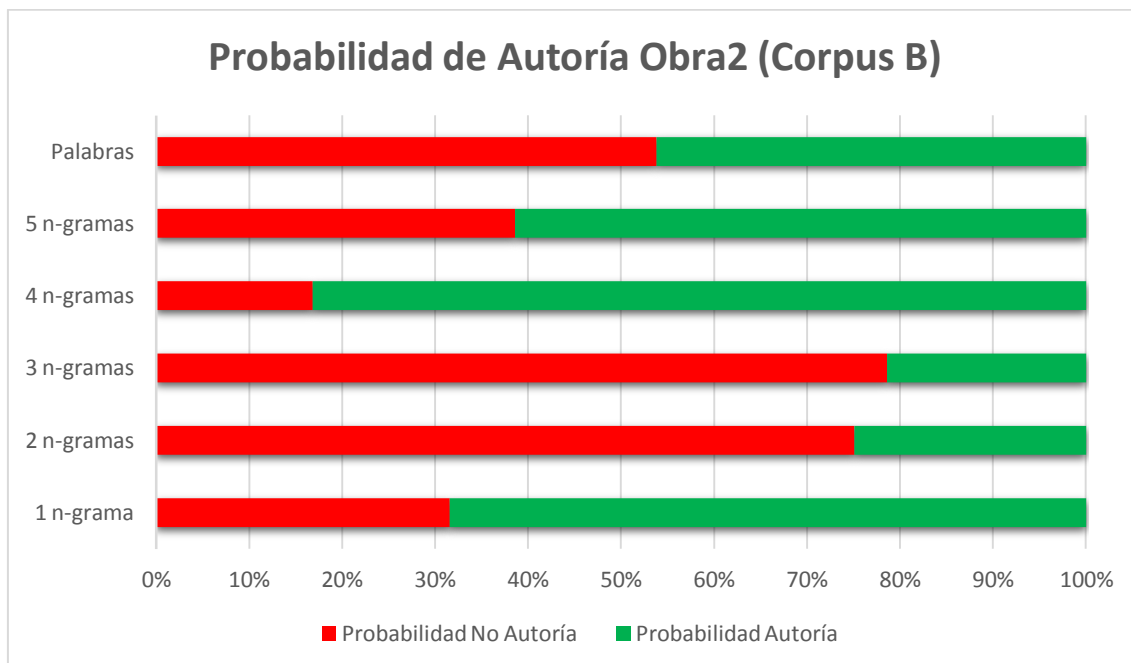
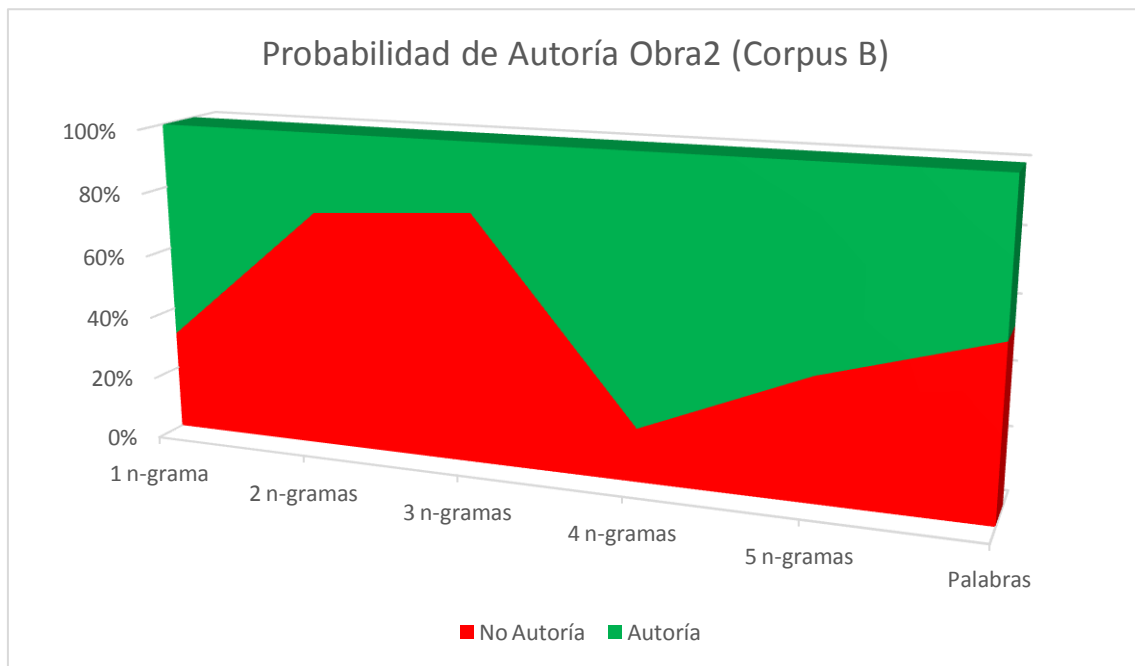


Gráfico 23



4.6.3. Análisis de la Obra 3 (Corpus B)

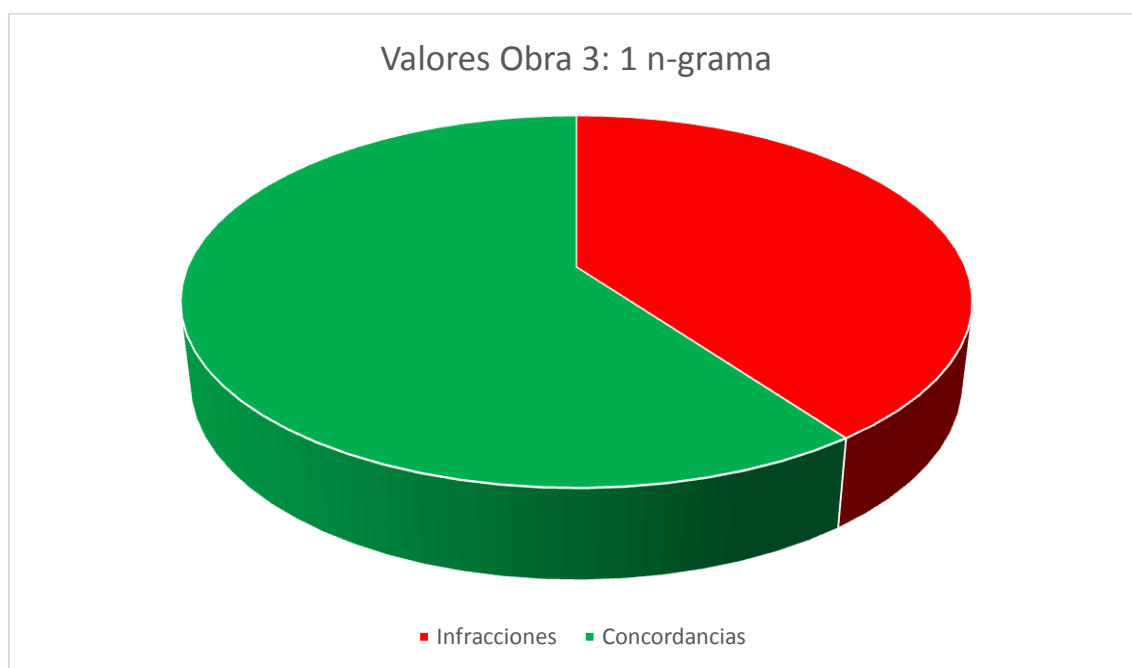
Título de la obra: *The Fire from Within*

Fecha de publicación: 1984

Tabla 35

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
“c”	1,93– 2,35	2,12		X
“s”	5,13 – 5,73	6,04	X	
“u”	2,76 – 3,18	2,62	X	
“f”	1,73 – 2,24	2,03		X
“j”	0,22 – 0,34	0,31		X
TOTAL			2	3

Gráfico 24

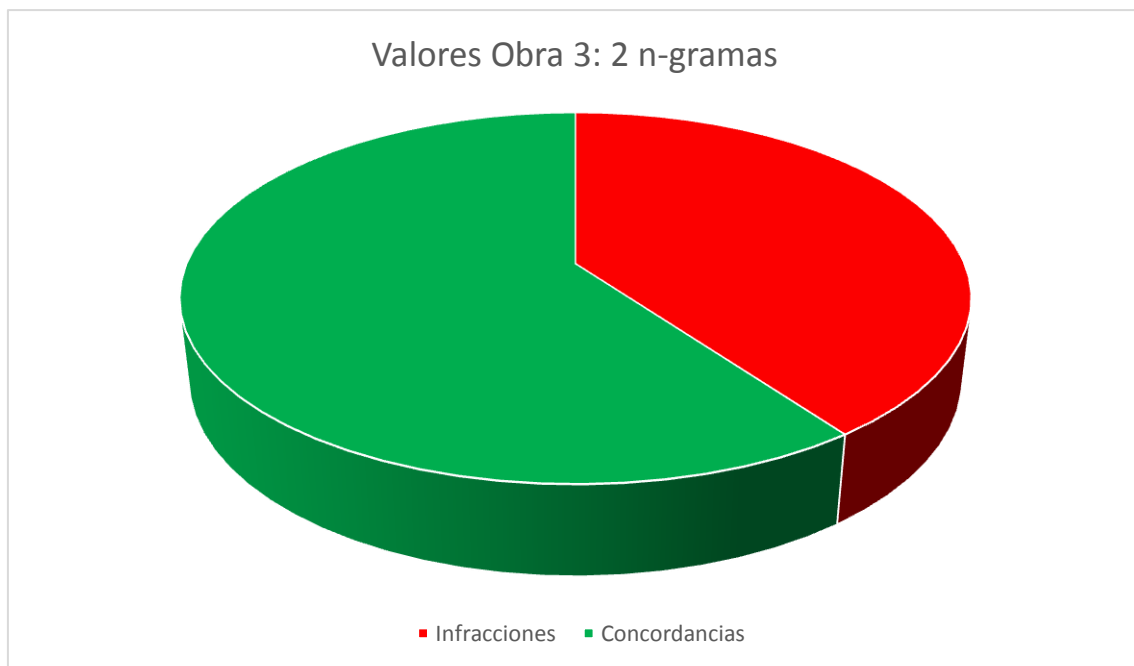


Infracciones y concordancias de 1 n-grama en la obra *The Fire from Within*

Tabla 36

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
er	1,31 – 1,43	1,45	X	
nd	0,92 – 1,13	0,83	X	
of	0,48 – 0,75	0,73		X
is	0,53 – 0,80	0,60		X
as	0,79 – 0,88	0,86		X
le	0,50 – 0,68	0,55		X
co	0,32 – 0,42	0,46	X	
ri	0,33 – 0,39	0,32	X	
ic	0,18 – 0,30	0,23		X
ur	0,28 – 0,40	0,31		X
TOTAL			4	6

Gráfico 25

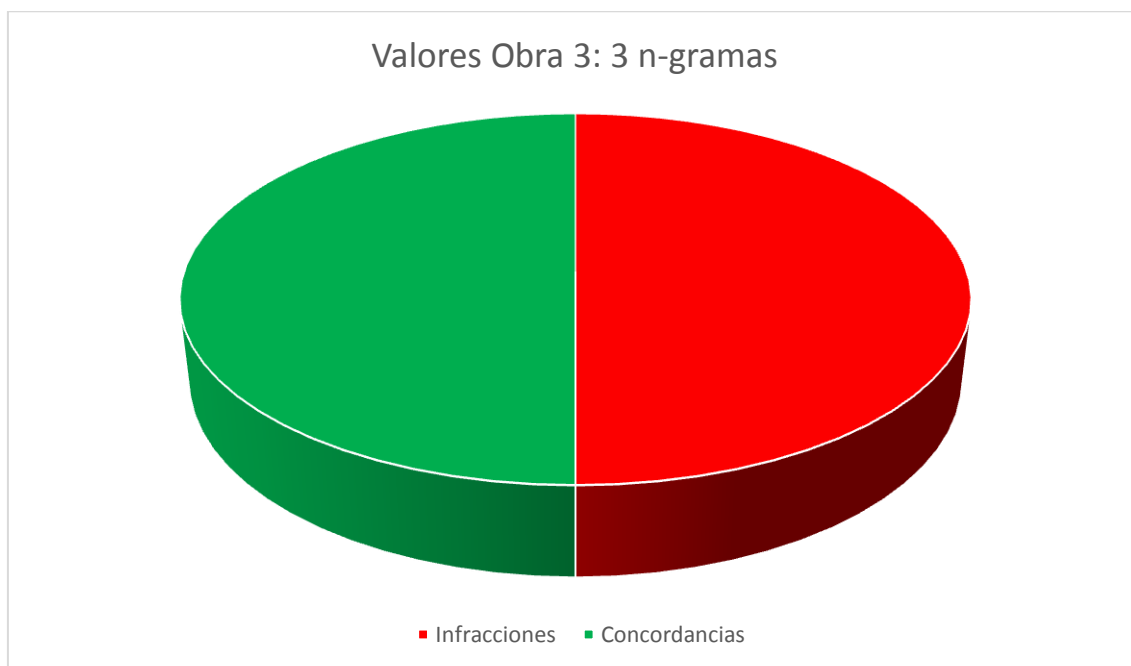


Infracciones y concordancias de 2 n-gramas en la obra *The Fire from Within*

Tabla 37

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
and	0,66 – 0,78	0,55	X	
ion	0,13 – 0,34	0,34		X
con	0,10 – 0,11	0,13	X	
ons	0,05 – 0,08	0,17	X	
nce	0,11 – 0,14	0,14		X
ive	0,05 – 0,11	0,08		X
was	0,39 – 0,48	0,33	X	
ica	0,02 – 0,05	0,04		X
cti	0,03 – 0,06	0,07	X	
rat	0,03 – 0,06	0,06		X
TOTAL			5	5

Gráfico 26

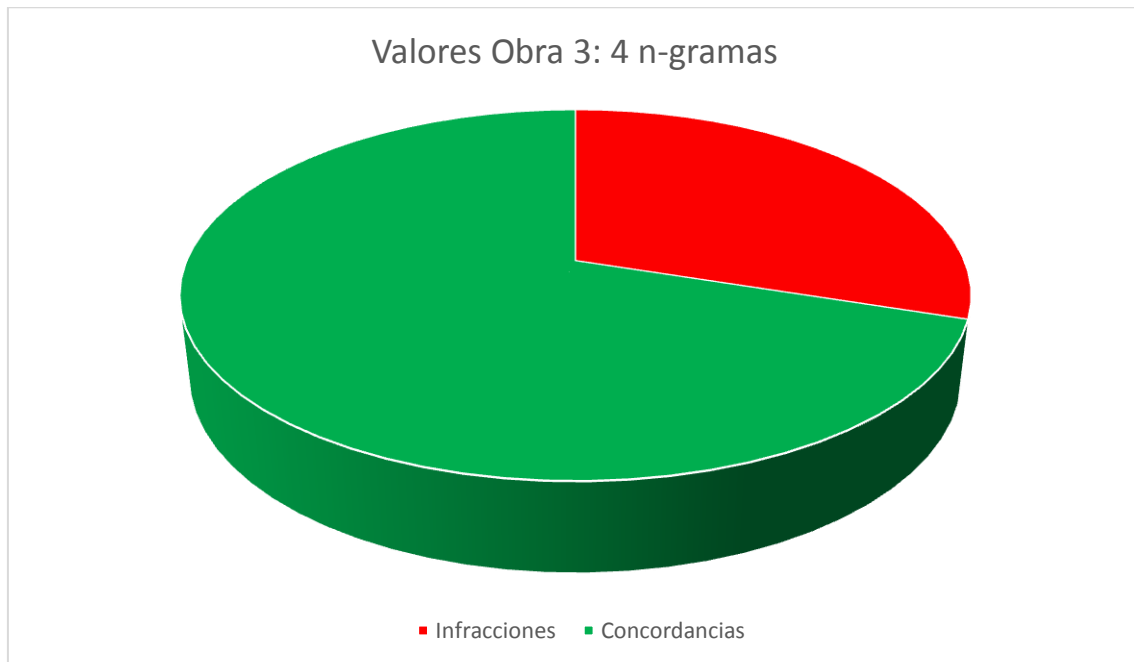


Infracciones y concordancias de 3 n-gramas en la obra *The Fire from Within*

Tabla 38

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
ions	0,02 – 0,05	0,12	X	
hich	0,02 – 0,05	0,03		X
ctio	0,02 – 0,04	0,03		X
ight	0,08 – 0,14	0,01	X	
sion	0,03 – 0,03	0,03		X
inte	0,02 – 0,05	0,05		X
nter	0,00 – 0,06	0,04		X
heir	0,00 – 0,05	0,10	X	
cons	0,00 – 0,02	0,02		X
рати	0,00 – 0,03	0,02		X
TOTAL			3	7

Gráfico 27

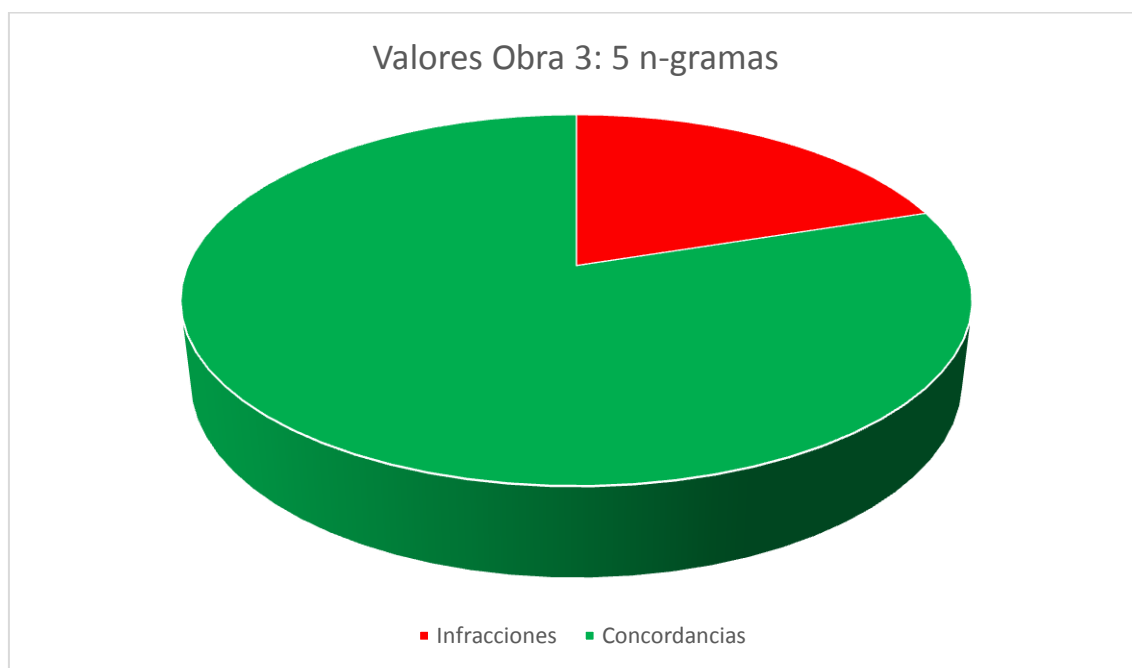


Infracciones y concordancias de 4 n-gramas en la obra *The Fire from Within*

Tabla 39

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
which	0,02 – 0,05	0,03		X
ction	0,03 – 0,03	0,03		X
ition	0,00 – 0,03	0,04	X	
ratio	0,00 – 0,02	0,02		X
state	0,01 – 0,01	0,02	X	
latio	0,00 – 0,01	0,00		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,00		X
gener	0,00 – 0,00	0,00		X
TOTAL			2	8

Gráfico 28

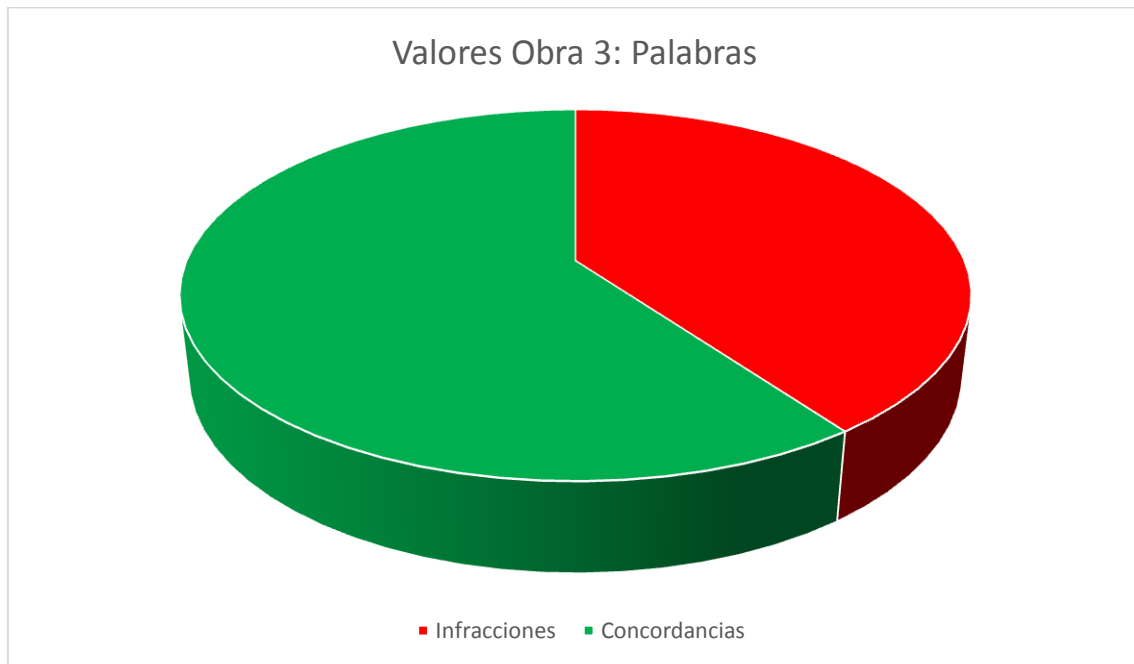


Infracciones y concordancias de 5 n-gramas en la obra *The Fire from Within*

Tabla 40

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O3	Infracciones	Concordancias
and	2,34 – 2,88	1,96	X	
with	0,56 – 0,77	0,70		X
be	0,36 – 0,51	0,37		X
by	0,19 – 0,46	0,41		X
this	0,03 – 0,33	0,17		X
are	0,09 – 0,45	0,54	X	
which	0,08 – 0,22	0,17		X
they	0,21 – 0,42	0,77	X	
you	1,00 – 1,87	0,61	X	
has	0,04 – 0,19	0,12		X
TOTAL			4	6

Gráfico 29



Infracciones y concordancias de palabras en la obra *The Fire from Within*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 3 del Corpus B.

Gráfico 30

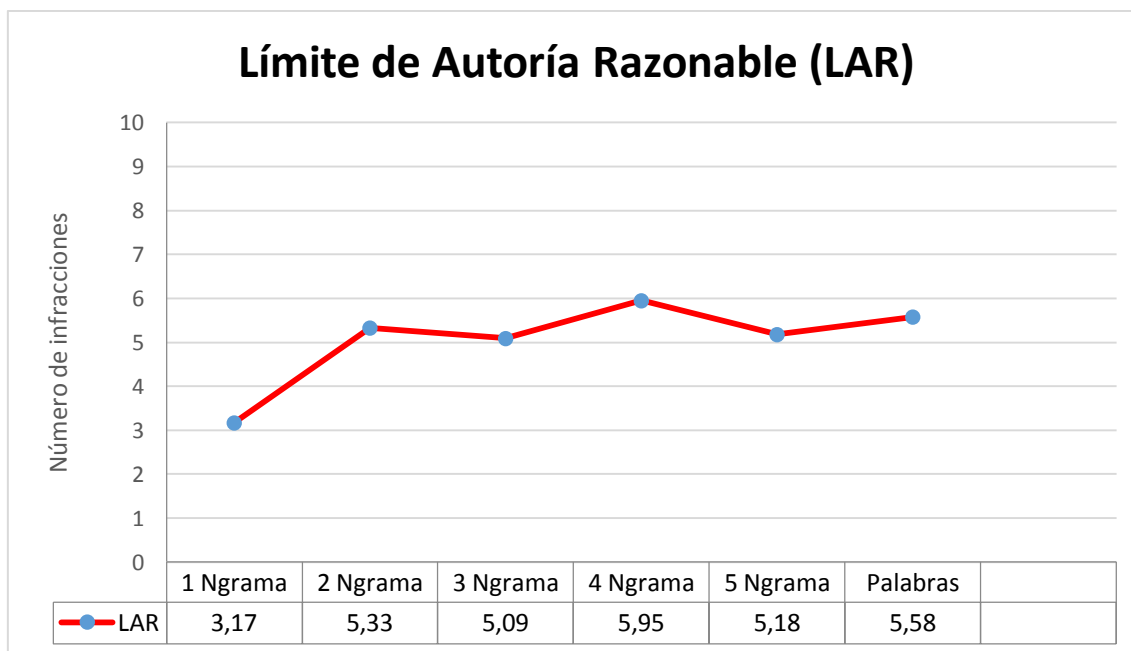


Gráfico 31

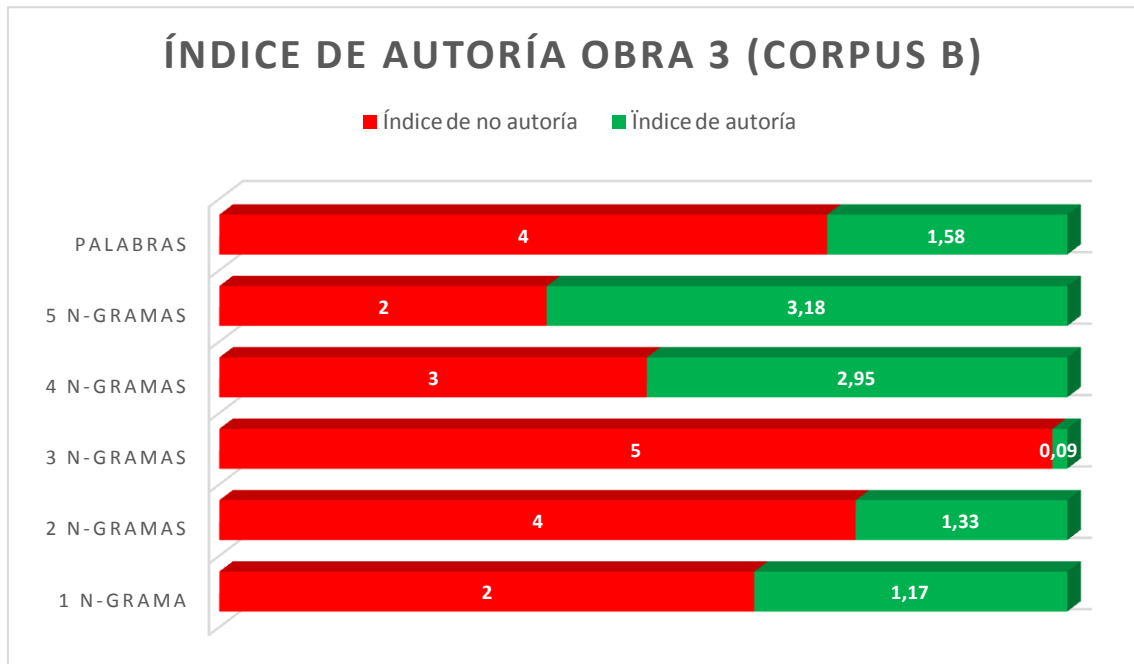


Gráfico 32

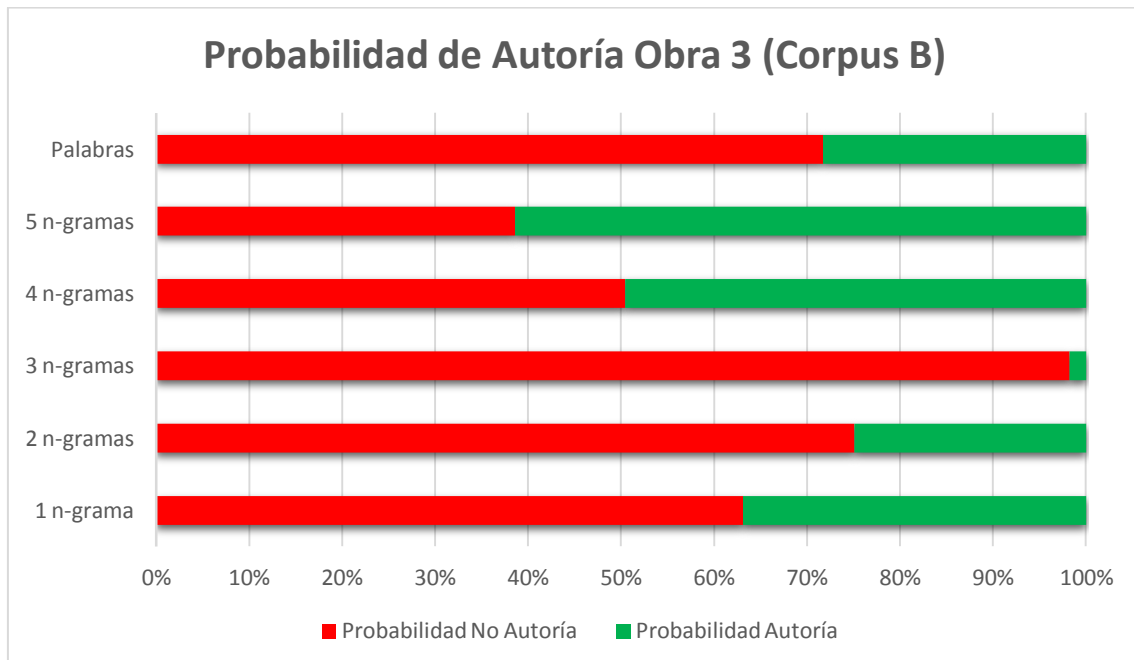
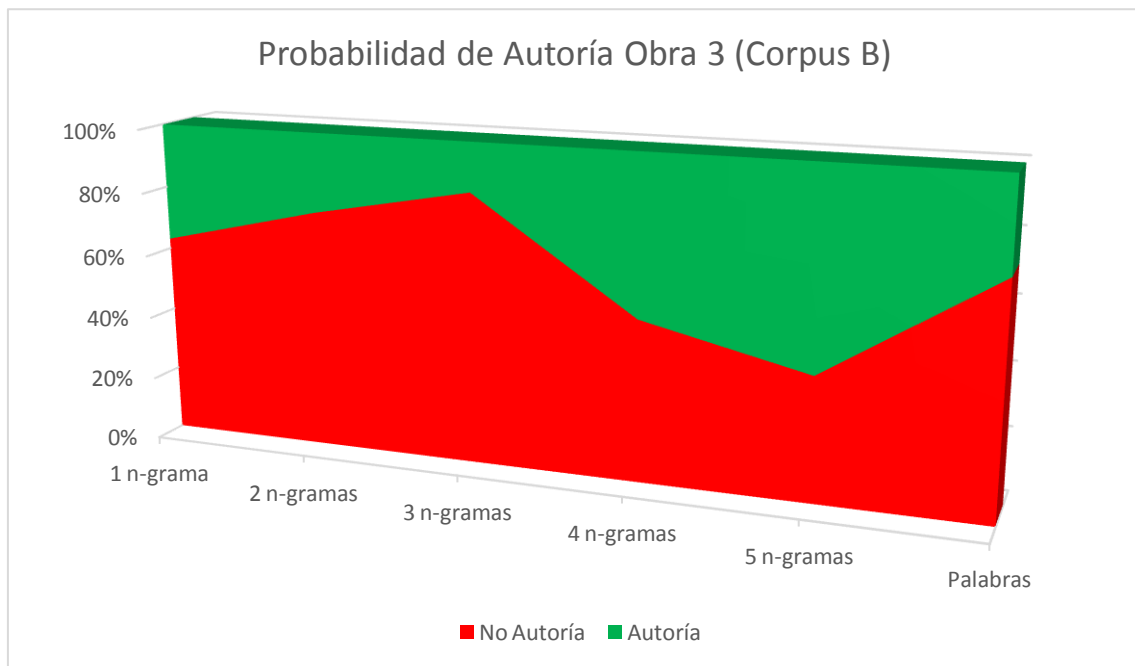


Gráfico 33



4.6.4. Análisis de la Obra 4 (Corpus B)

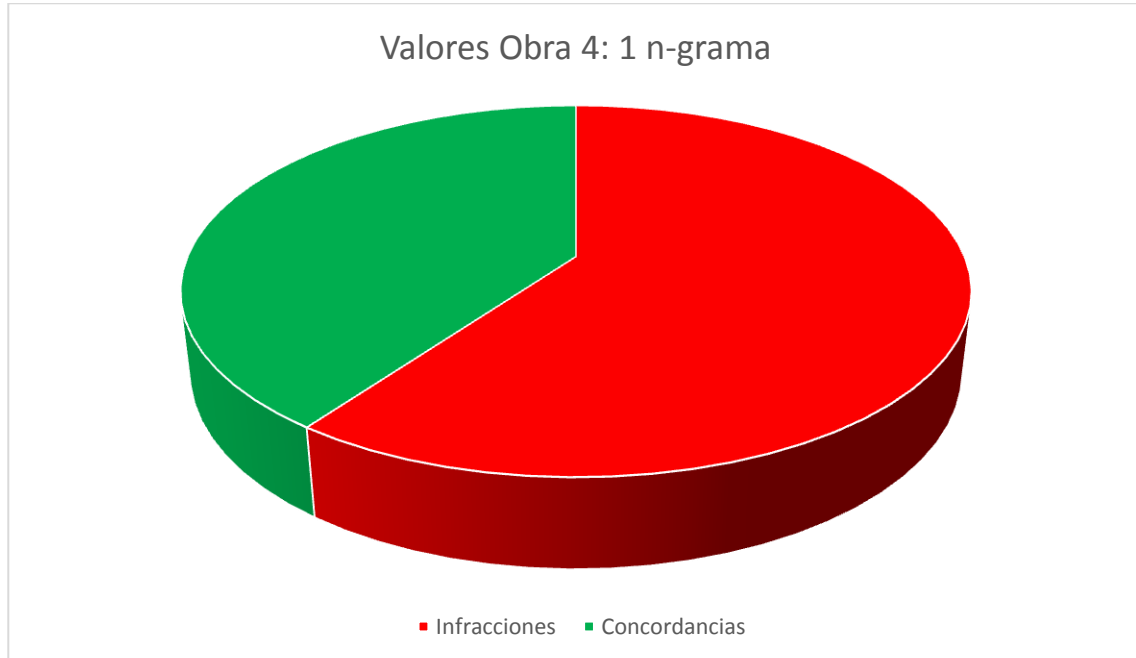
Título de la obra: *The Power of Silence: Further Lessons of Don Juan*

Fecha de publicación: 1987

Tabla 41

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
“c”	1,93 – 2,35	2,50	X	
“s”	5,13 – 5,73	5,91	X	
“u”	2,76 – 3,18	2,94		X
“f”	1,73 – 2,24	1,96		X
“j”	0,22 – 0,34	0,37	X	
TOTAL			3	2

Gráfico 34

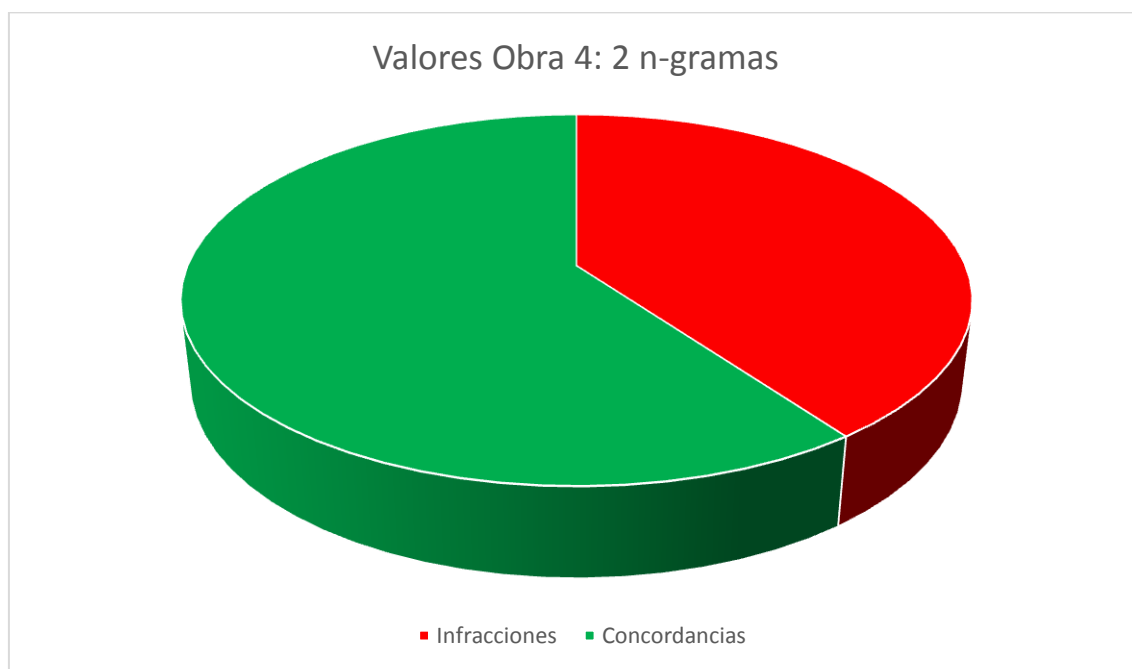


Infracciones y concordancias de 1 n-grama en la obra *The Power of Silence: Further Lessons of Don Juan*

Tabla 42

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
er	1,31 – 1,43	1,49	X	
nd	0,92 – 1,13	0,87	X	
of	0,48 – 0,75	0,67		X
is	0,53 – 0,80	0,70		X
as	0,79 – 0,88	0,96	X	
le	0,50 – 0,68	0,58		X
co	0,32 – 0,42	0,48	X	
ri	0,33 – 0,39	0,39		X
ic	0,18 – 0,30	0,30		X
ur	0,28 – 0,40	0,32		X
TOTAL			4	6

Gráfico 35

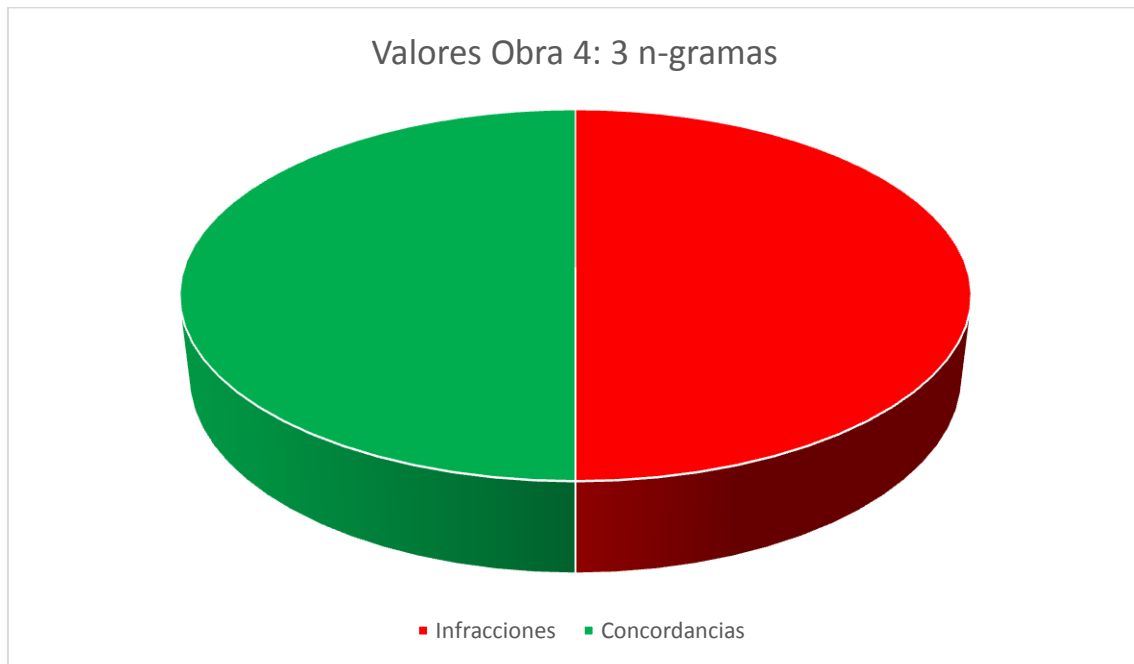


Infracciones y concordancias de 2 n-gramas en la obra *The Power of Silence: Further Lessons of Don Juan*

Tabla 43

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
and	0,66 – 0,78	0,56	X	
ion	0,13 – 0,34	0,29		X
con	0,10 – 0,11	0,17	X	
ons	0,05 – 0,08	0,12	X	
nce	0,11 – 0,14	0,17	X	
ive	0,05 – 0,11	0,10		X
was	0,39 – 0,48	0,45		X
ica	0,02 – 0,05	0,05		X
cti	0,03 – 0,06	0,10	X	
rat	0,03 – 0,06	0,06		X
TOTAL			5	5

Gráfico 36

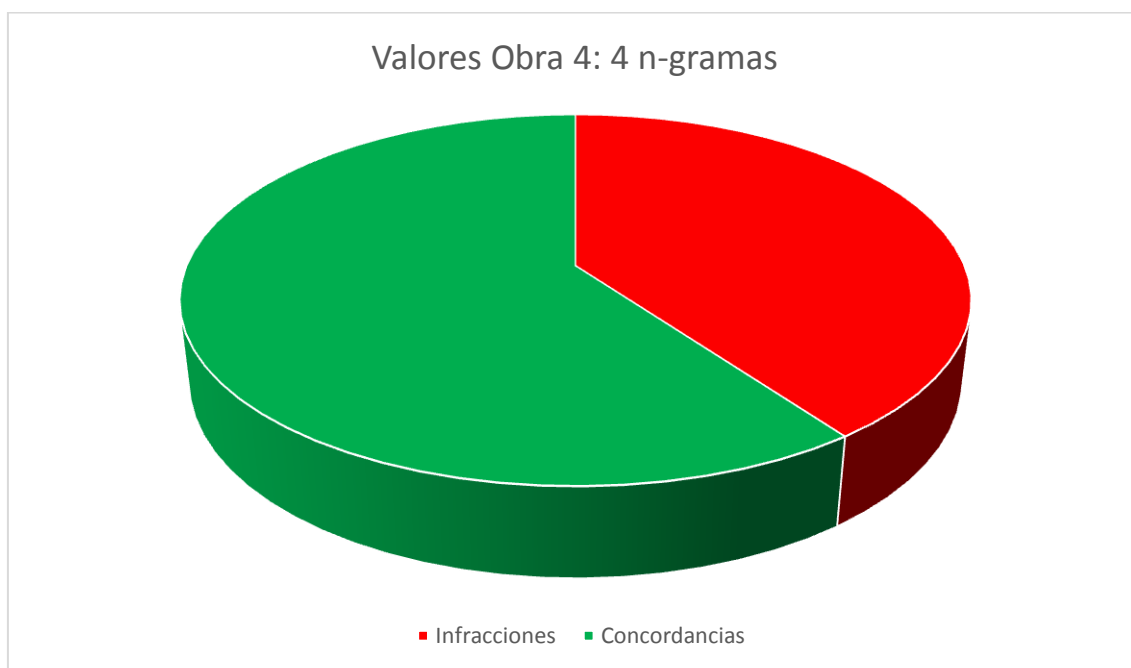


Infracciones y concordancias de 3 n-gramas en la obra *The Power of Silence: Further Lessons of Don Juan*

Tabla 44

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
ions	0,02 – 0,05	0,05		X
hich	0,02 – 0,05	0,05		X
ctio	0,02 – 0,04	0,06	X	
ight	0,08 – 0,14	0,09		X
sion	0,03 – 0,03	0,03		X
inte	0,02 – 0,05	0,07	X	
nter	0,00 – 0,06	0,03		X
heir	0,00 – 0,05	0,06	X	
cons	0,00 – 0,02	0,03	X	
рати	0,00 – 0,03	0,03		X
TOTAL			4	6

Gráfico 37

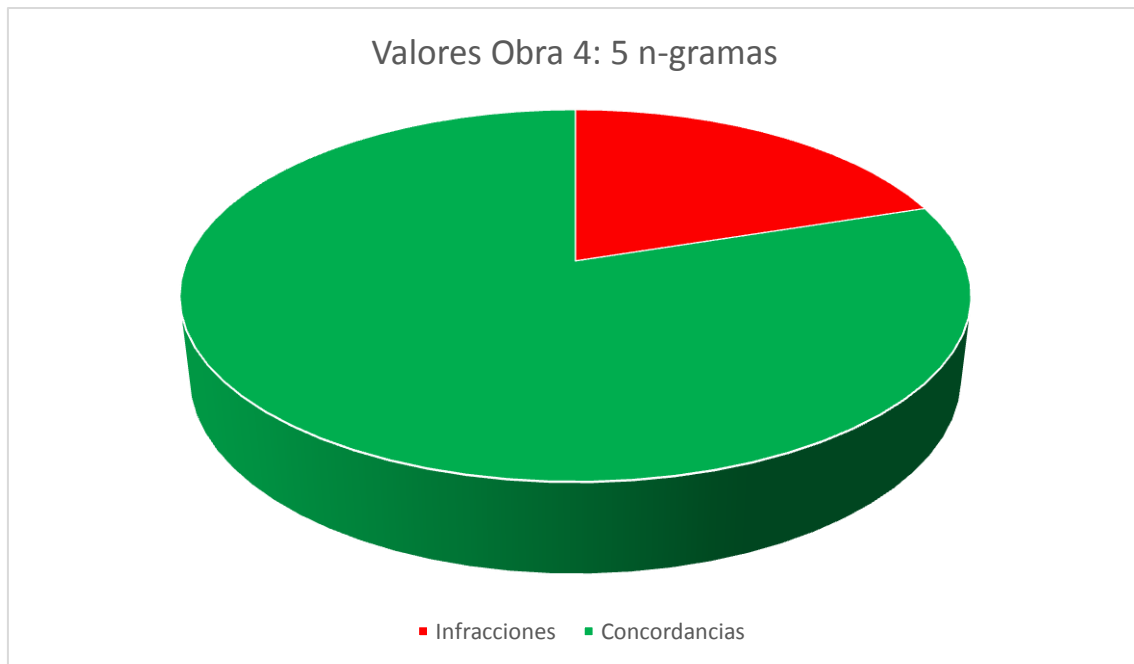


Infracciones y concordancias de 4 n-gramas en la obra *The Power of Silence: Further Lessons of Don Juan*

Tabla 45

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
which	0,02 – 0,05	0,05		X
ction	0,03 – 0,03	0,06	X	
ition	0,00 – 0,03	0,03		X
ratio	0,00 – 0,02	0,02		X
state	0,01 – 0,01	0,02	X	
latio	0,00 – 0,01	0,00		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,01		X
gener	0,000–0,00	0,00		X
TOTAL			2	8

Gráfico 38

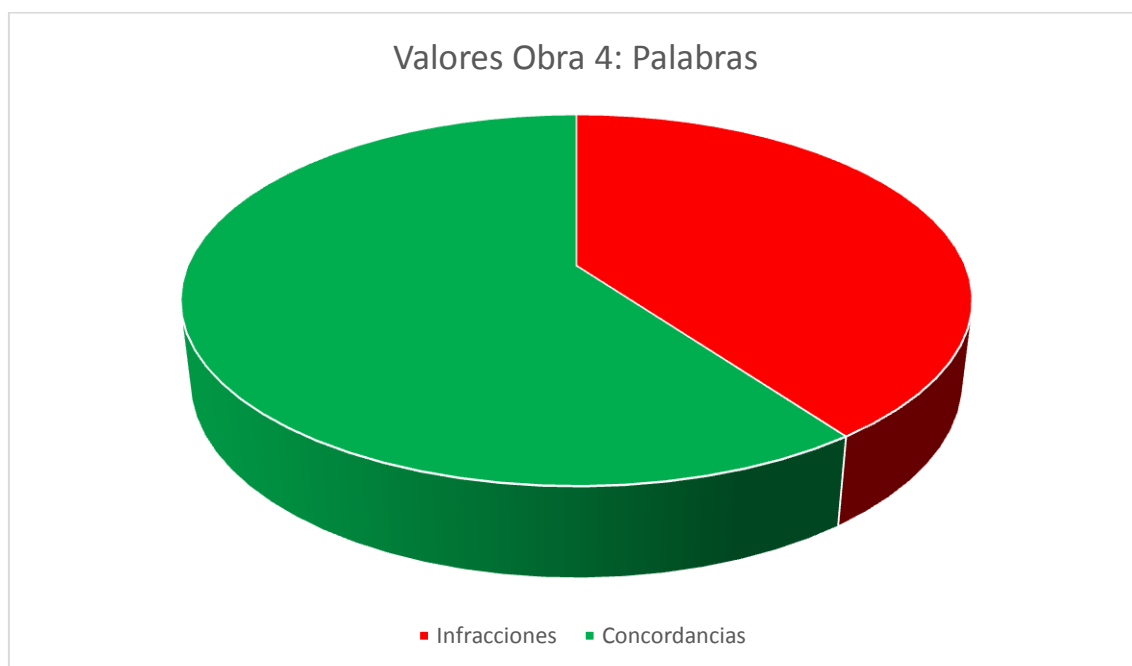


Infracciones y concordancias de 5 n-gramas en la obra *The Power of Silence: Further Lessons of Don Juan*

Tabla 46

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O4	Infracciones	Concordancias
and	2,34 – 2,88	2,23	X	
with	0,56 – 0,77	0,69		X
be	0,36 – 0,51	0,41		X
by	0,19 – 0,46	0,41		X
this	0,03 – 0,33	0,28		X
are	0,09 – 0,45	0,30		X
which	0,08 – 0,22	0,24	X	
they	0,21 – 0,42	0,46	X	
you	1,00 – 1,87	0,71	X	
has	0,04 – 0,19	0,08		X
TOTAL			4	6

Gráfico 39



Infracciones y concordancias de palabras en la obra *The Power of Silence: Further Lessons of Don Juan*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 4 del Corpus B.

Gráfico 40

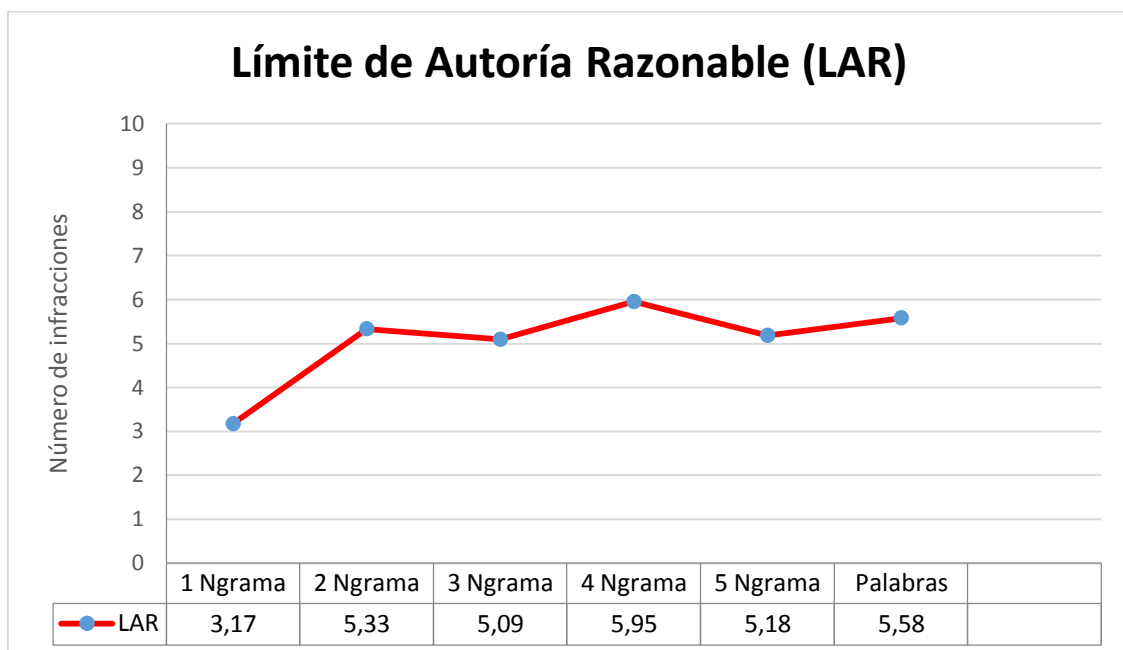


Gráfico 41

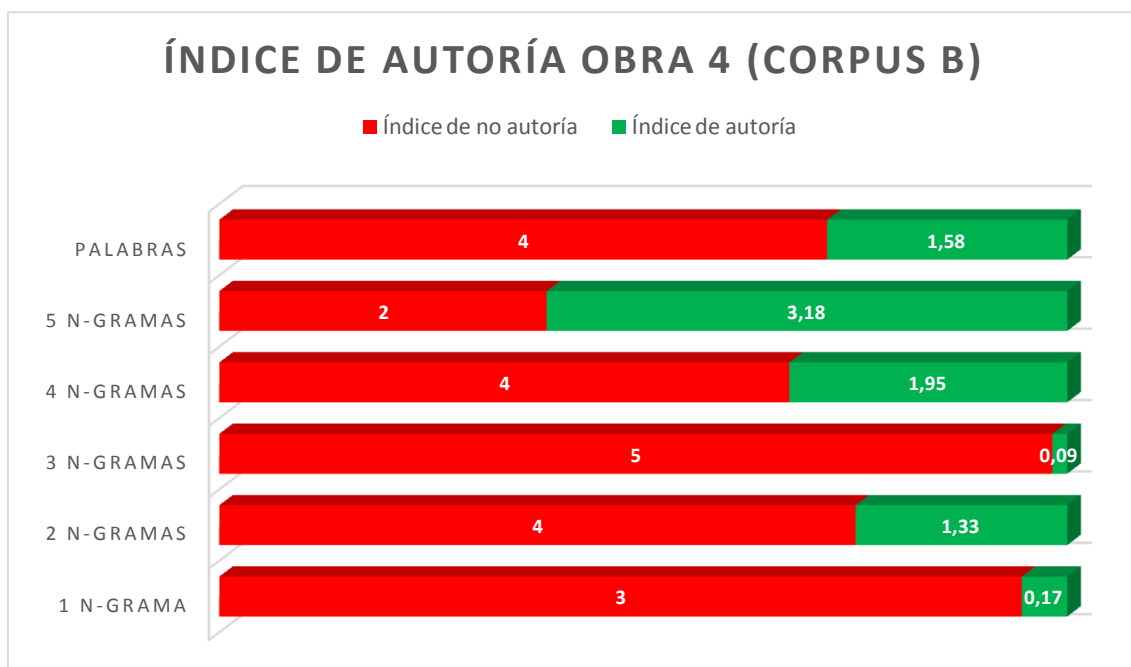


Gráfico 42

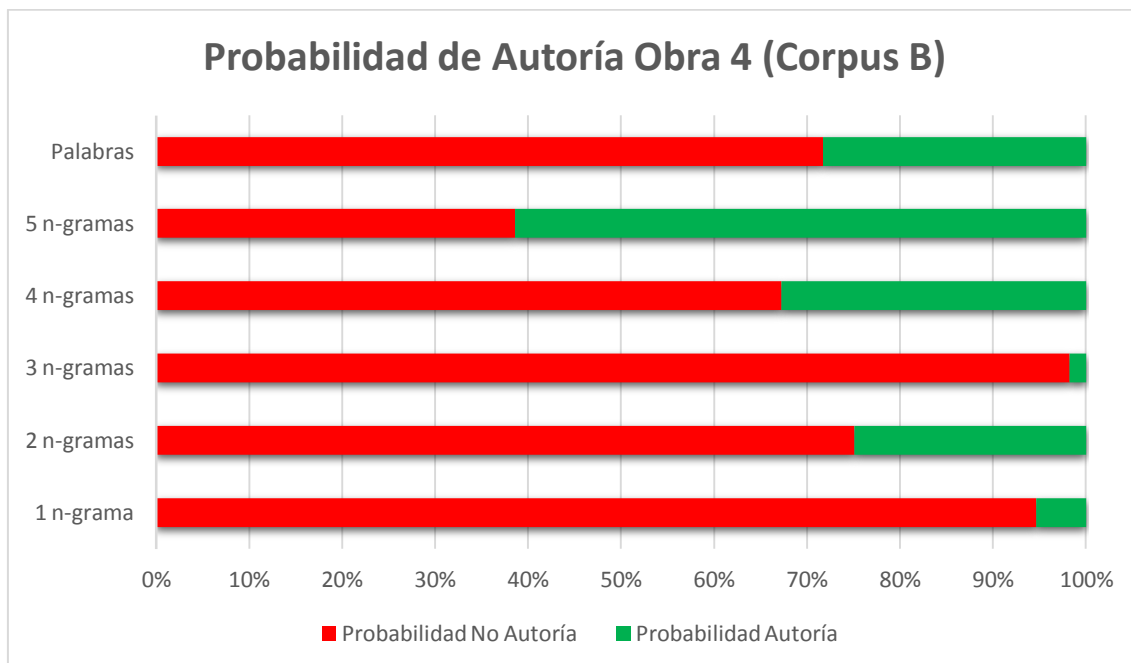
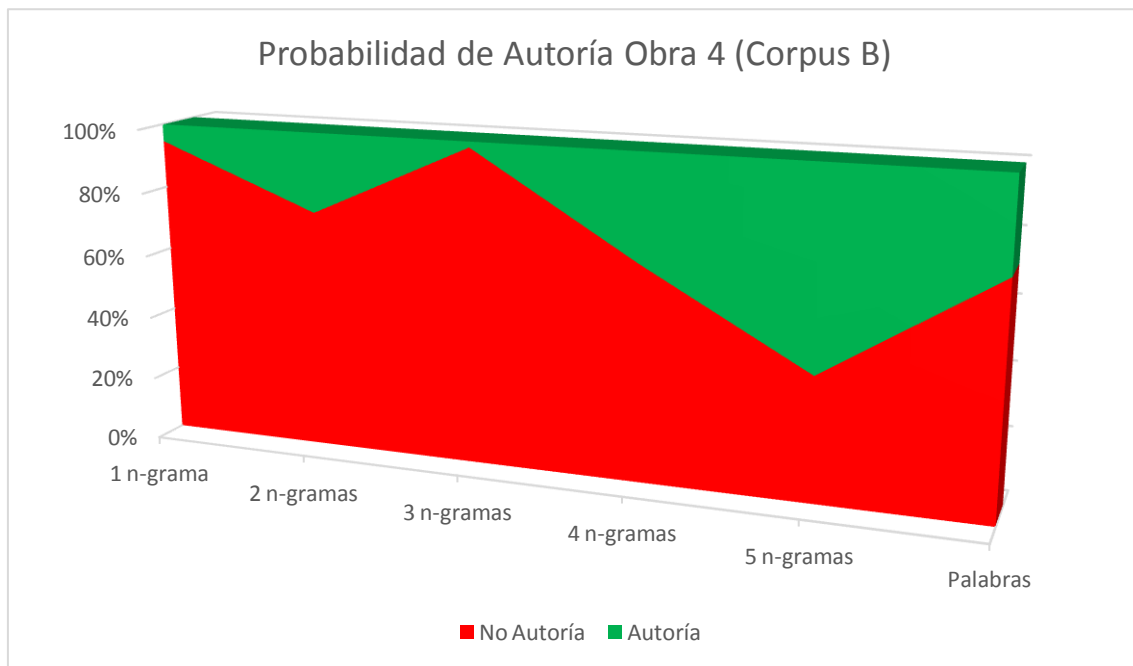


Gráfico 43



4.6.5. Análisis de la Obra 5 (Corpus B)

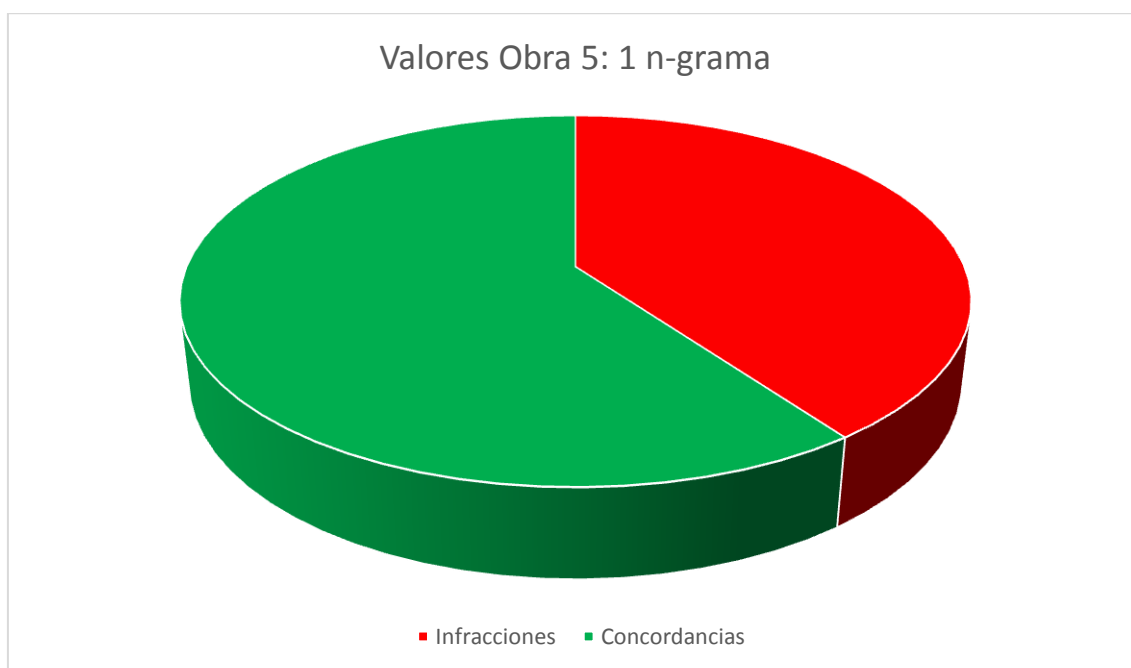
Título de la obra: *The Art of Dreaming*

Fecha de publicación: 1993

Tabla 47

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
“c”	1,93 – 2,35	2,36	X	
“s”	5,13 – 5,73	5,80	X	
“u”	2,76 – 3,18	2,86		X
“f”	1,73 – 2,24	1,88		X
“j”	0,22 – 0,34	0,27		X
TOTAL			2	3

Gráfico 44

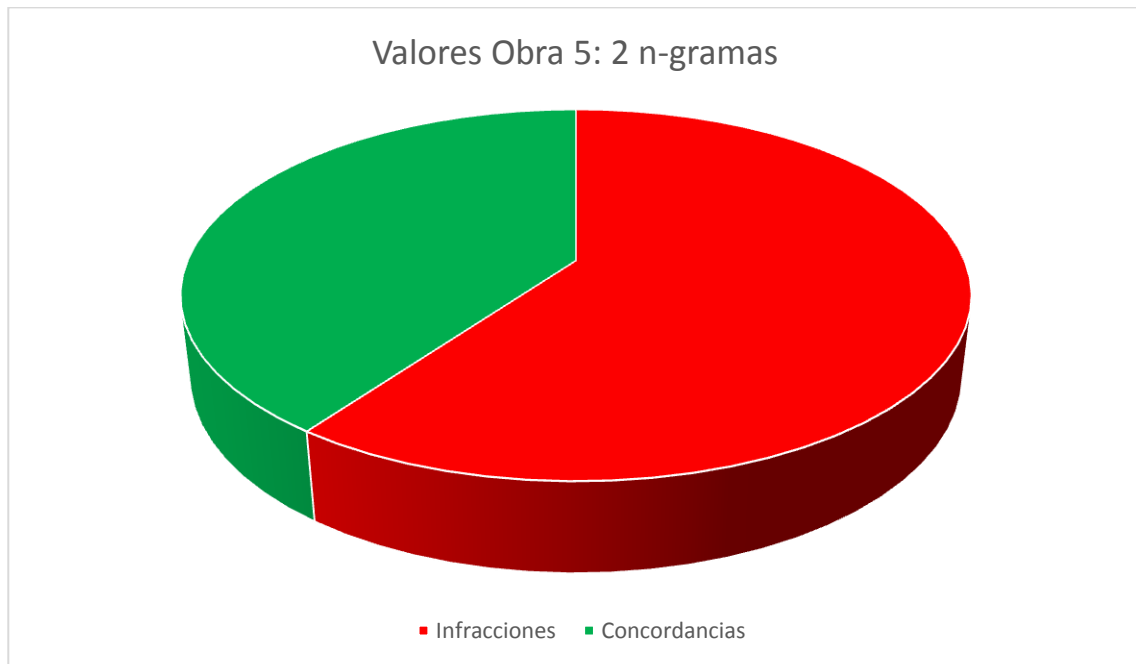


Infracciones y concordancias de 1 n-grama en la obra *The Art of Dreaming*

Tabla 48

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
er	1,31 – 1,43	1,64	X	
nd	0,92 – 1,13	0,80	X	
of	0,48 – 0,75	0,67		X
is	0,53 – 0,80	0,67		X
as	0,79 – 0,88	0,81		X
le	0,50 – 0,68	0,45	X	
co	0,32 – 0,42	0,44	X	
ri	0,33 – 0,39	0,30	X	
ic	0,18 – 0,30	0,35	X	
ur	0,28 – 0,40	0,33		X
TOTAL			6	4

Gráfico 45

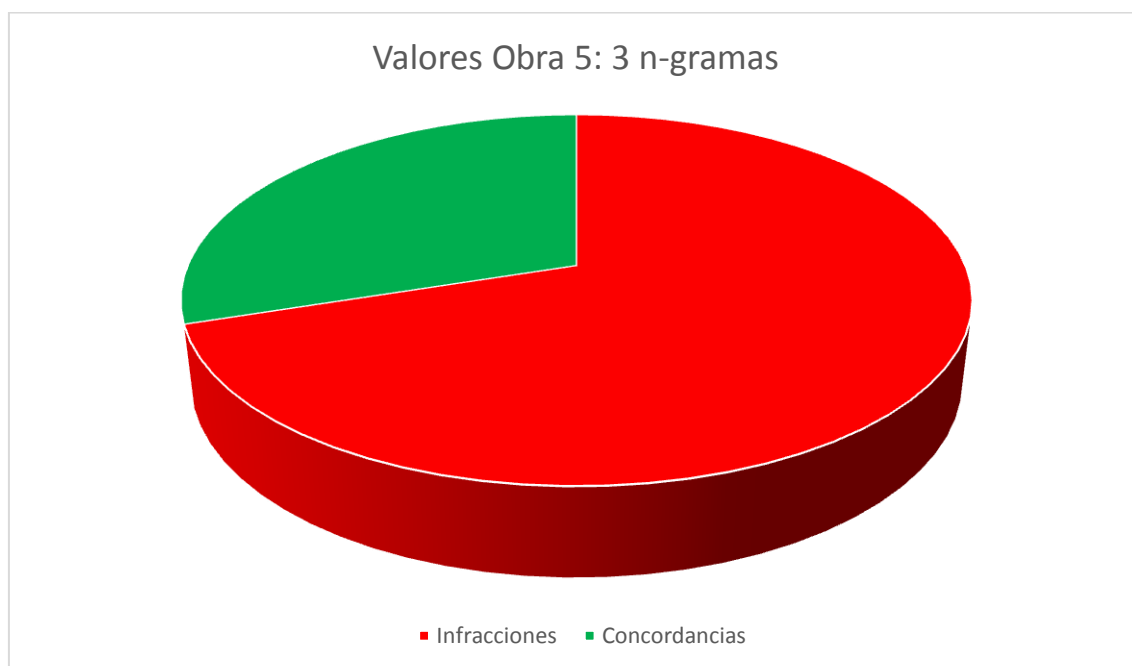


Infracciones y concordancias de 2 n-gramas en la obra *The Art of Dreaming*

Tabla 49

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
and	0,66 – 0,78	0,45	X	
ion	0,13 – 0,34	0,36	X	
con	0,10 – 0,11	0,17	X	
ons	0,05 – 0,08	0,12	X	
nce	0,11 – 0,14	0,14		X
ive	0,05 – 0,11	0,12	X	
was	0,39 – 0,48	0,34	X	
ica	0,02 – 0,05	0,05		X
cti	0,03 – 0,06	0,08	X	
rat	0,03 – 0,06	0,05		X
TOTAL			7	3

Gráfico 46

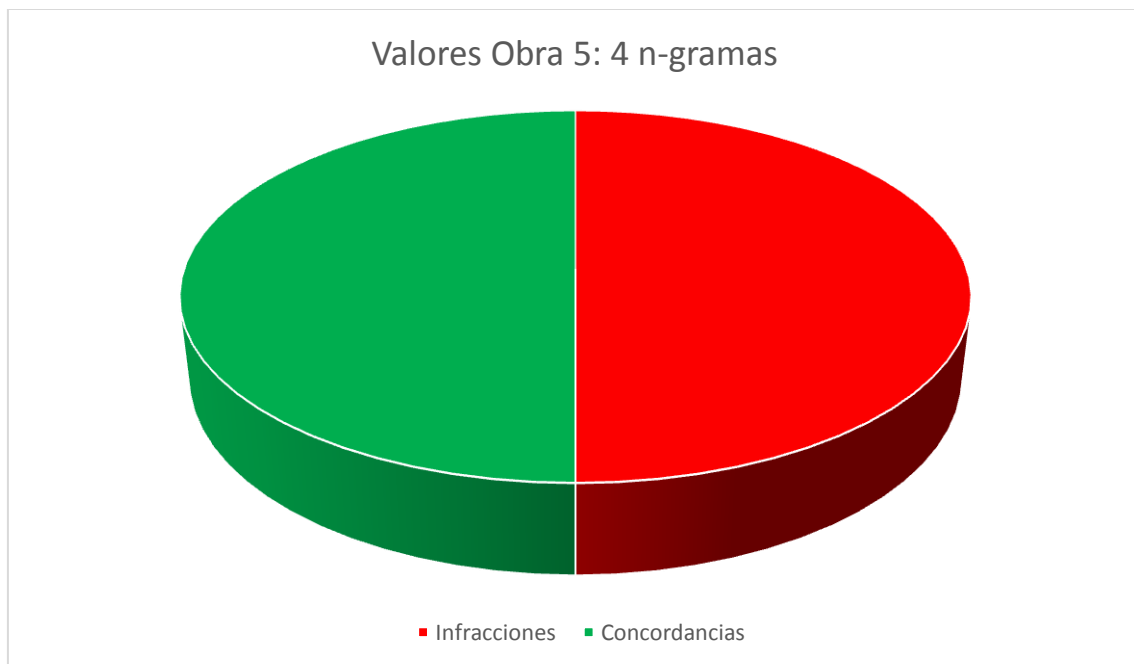


Infracciones y concordancias de 3 n-gramas en la obra *The Art of Dreaming*

Tabla 50

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
ions	0,02 – 0,05	0,07	X	
hich	0,02 – 0,05	0,04		X
ctio	0,02 – 0,04	0,03		X
ight	0,08 – 0,14	0,08		X
sion	0,03 – 0,03	0,04	X	
inte	0,02 – 0,05	0,07	X	
nter	0,00 – 0,06	0,04		X
heir	0,00 – 0,05	0,06	X	
cons	0,00 – 0,02	0,03	X	
рати	0,00 – 0,03	0,03		X
TOTAL			5	5

Gráfico 47

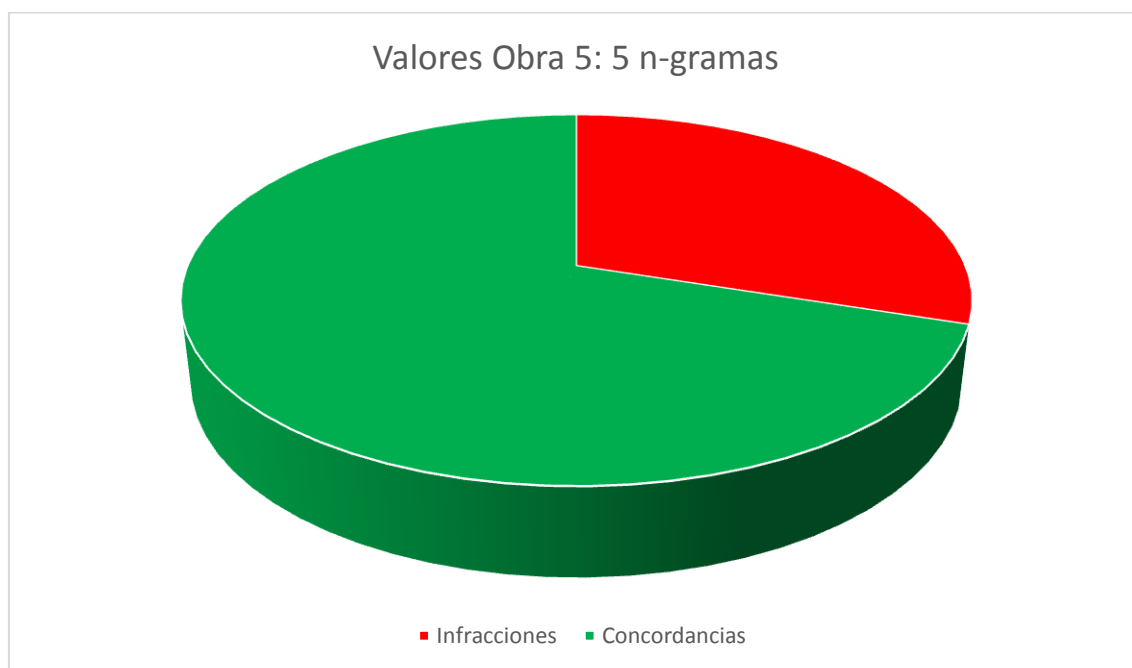


Infracciones y concordancias de 4 n-gramas en la obra *The Art of Dreaming*

Tabla 51

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
which	0,02 – 0,05	0,04		X
ction	0,03 – 0,03	0,03		X
ition	0,00 – 0,03	0,04	X	
ratio	0,00 – 0,02	0,01		X
state	0,01 – 0,01	0,02	X	
latio	0,00 – 0,01	0,01		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,01		X
gener	0,00 – 0,00	0,01	X	
TOTAL			3	7

Gráfico 48

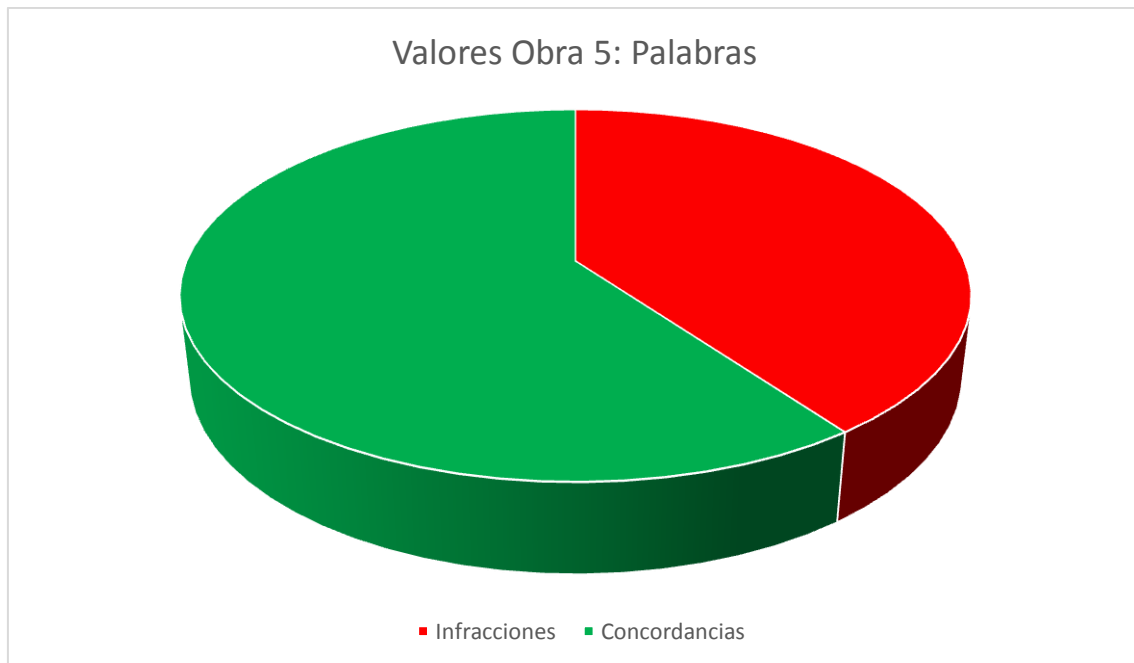


Infracciones y concordancias de 5 n-gramas en la obra *The Art of Dreaming*

Tabla 52

Categoría de marcador: palabras				
Marcador	Valor RdA	Valores O5	Infracciones	Concordancias
and	2,34 – 2,88	1,80	X	
with	0,56 – 0,77	0,65		X
be	0,36 – 0,51	0,39		X
by	0,19 – 0,46	0,42		X
this	0,03 – 0,33	0,49	X	
are	0,09 – 0,45	0,56	X	
which	0,08 – 0,22	0,19		X
they	0,21 – 0,42	0,52	X	
you	1,00 – 1,87	1,40		X
has	0,04 – 0,19	0,11		X
TOTAL			4	6

Gráfico 49



Infracciones y concordancias de palabras en la obra *The Art of Dreaming*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 5 del Corpus B.

Gráfico 50

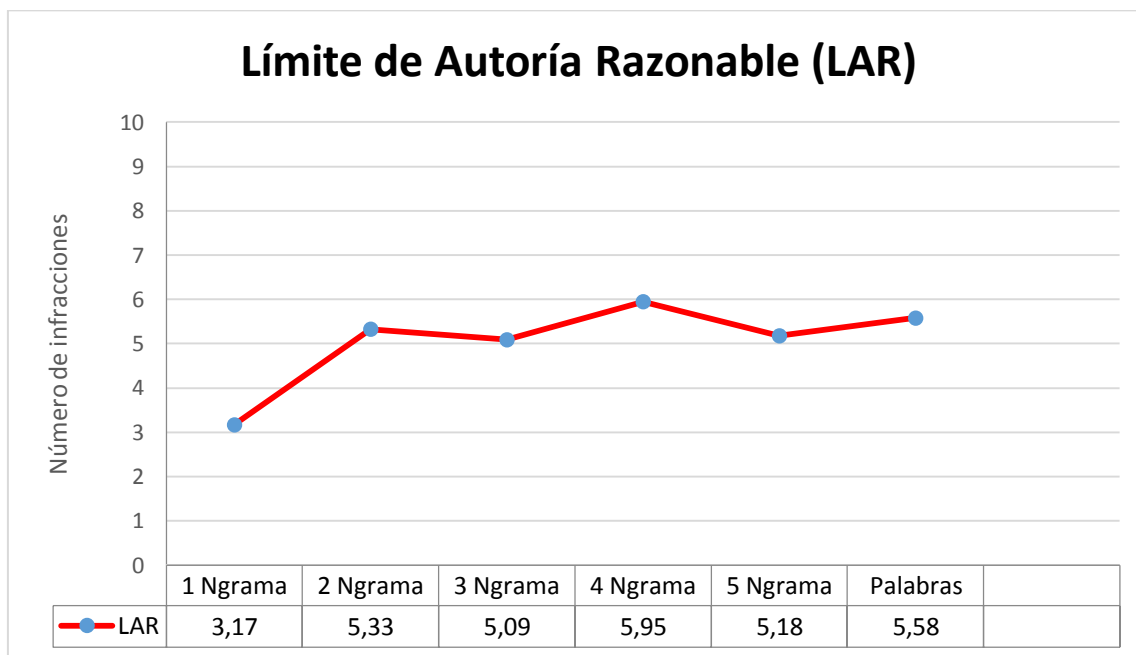


Gráfico 51

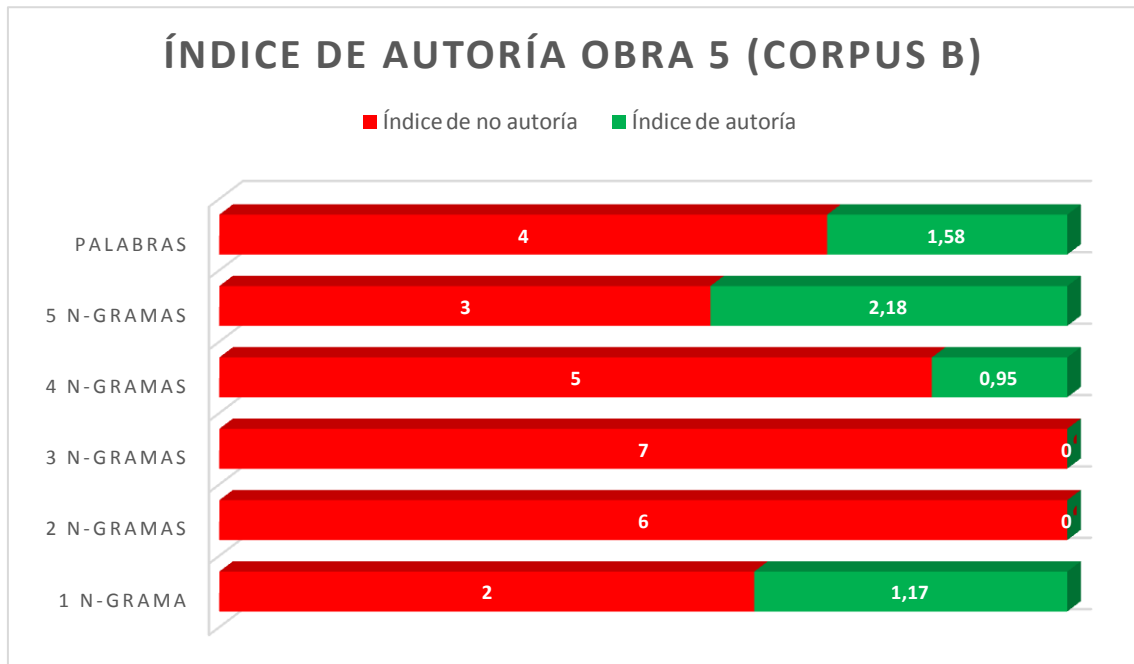


Gráfico 52

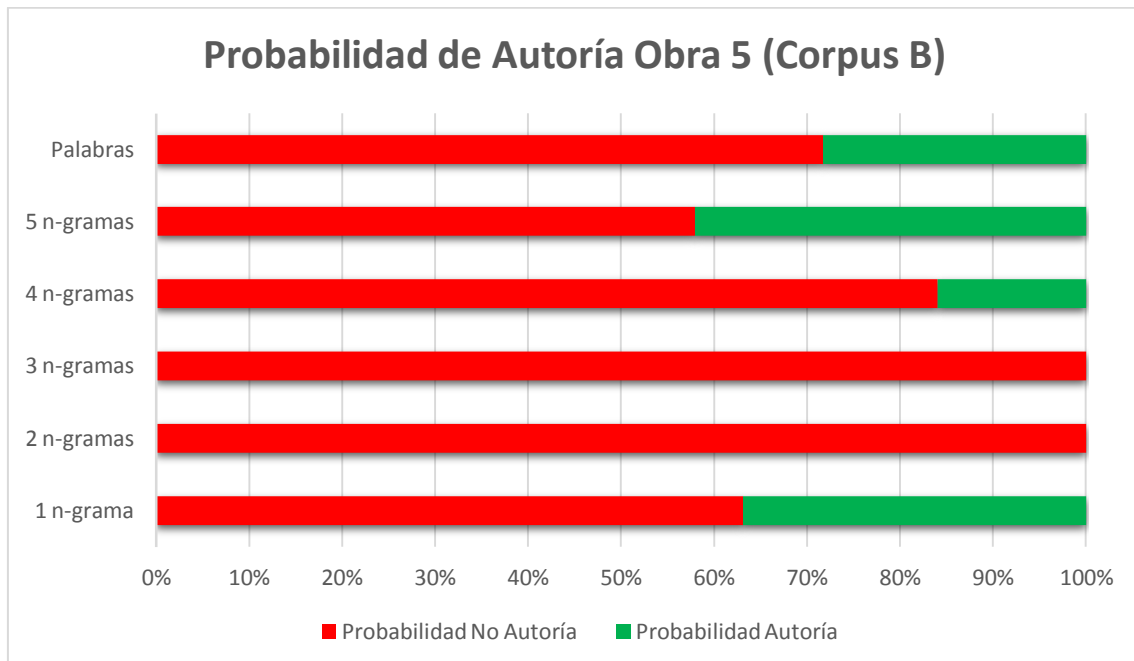
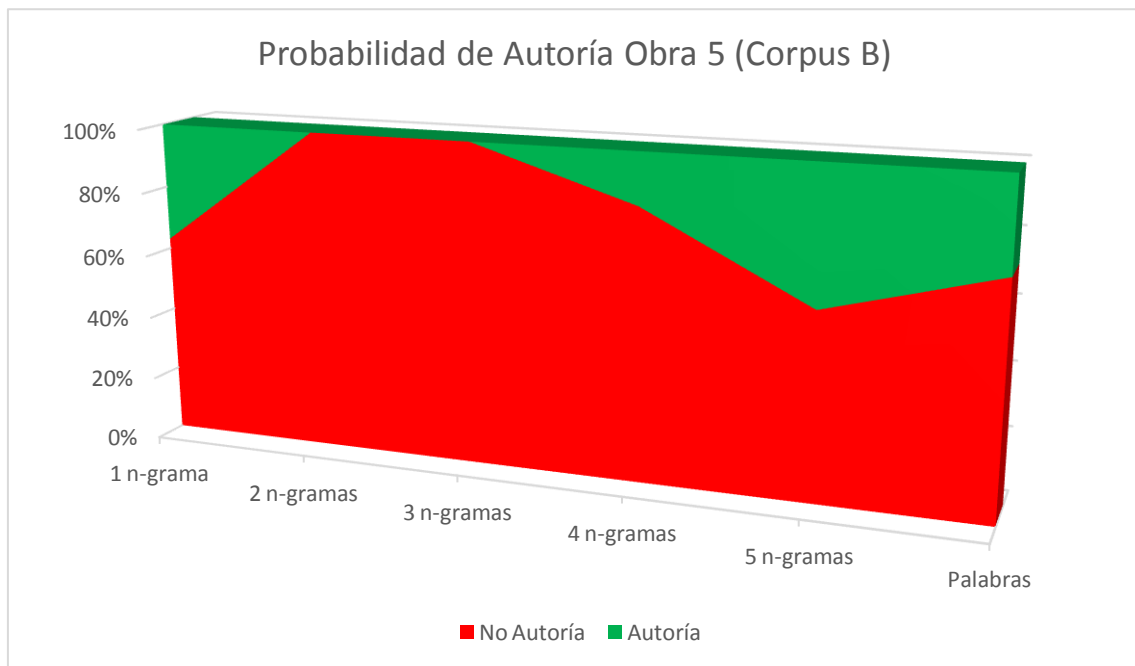


Gráfico 53



4.6.6. Análisis de la Obra 6 (Corpus B)

Título de la obra: *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Fecha de publicación: 1998

Tabla 53

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
“c”	1,93 – 2,35	2,98	X	
“s”	5,13 – 5,73	6,37	X	
“u”	2,76 – 3,18	2,29	X	
“p”	1,73 – 2,24	2,52	X	
“j”	0,22 – 0,34	0,21	X	
TOTAL			5	0

Gráfico 54

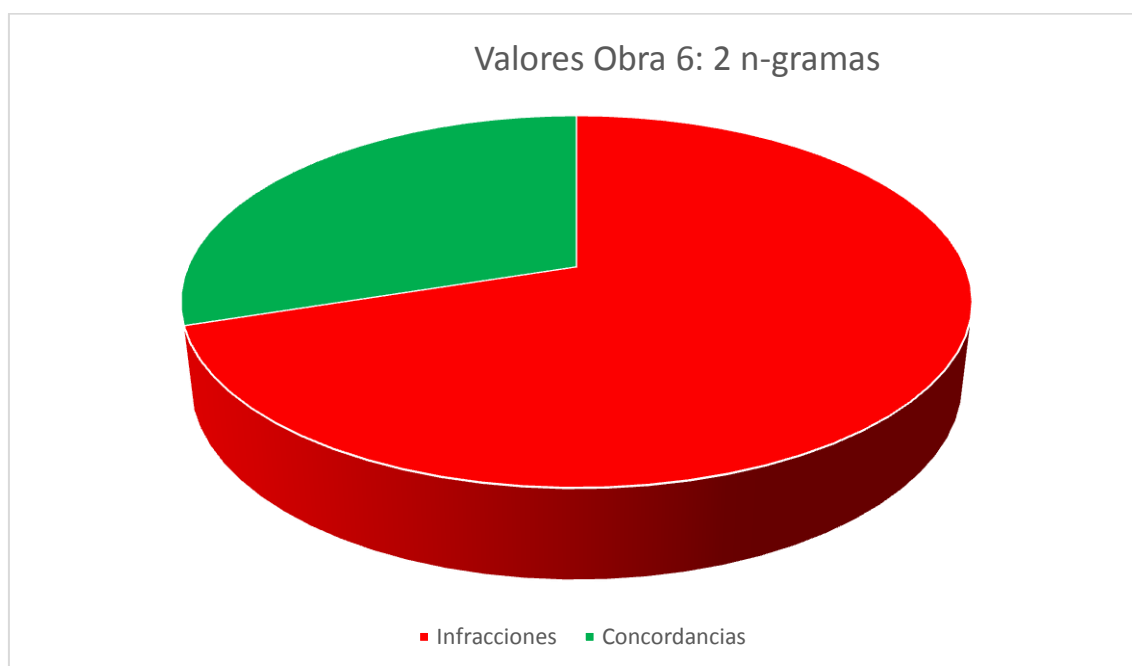


Infracciones y concordancias de 1 n-grama en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Tabla 54

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
er	1,31 – 1,43	1,76	X	
nd	0,92 – 1,13	0,84	X	
of	0,48 – 0,75	1,11	X	
is	0,53 – 0,80	0,87	X	
as	0,79 – 0,88	0,78	X	
le	0,50 – 0,68	0,65		X
co	0,32 – 0,42	0,42		X
ri	0,33 – 0,39	0,44	X	
ic	0,18 – 0,30	0,57	X	
ur	0,28 – 0,40	0,28		X
TOTAL			7	3

Gráfico 55



Infracciones y concordancias de 2 n-gramas en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Tabla 55

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
and	0,66 – 0,78	0,50	X	
ion	0,13 – 0,34	0,52	X	
con	0,10 – 0,11	0,19	X	
ons	0,05 – 0,08	0,13	X	
nce	0,11 – 0,14	0,18	X	
ive	0,05 – 0,11	0,16	X	
was	0,39 – 0,48	0,17	X	
ica	0,02 – 0,05	0,20	X	
cti	0,03 – 0,06	0,14	X	
rat	0,03 – 0,06	0,10	X	
TOTAL			10	0

Gráfico 56



Infracciones y concordancias de 3 n-gramas en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Tabla 56

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
ions	0,02 – 0,05	0,06	X	
hich	0,02 – 0,05	0,13	X	
ctio	0,02 – 0,04	0,05	X	
ight	0,08 – 0,14	0,14		X
sion	0,03 – 0,03	0,06	X	
inte	0,02 – 0,05	0,09	X	
nter	0,00 – 0,06	0,11	X	
heir	0,00 – 0,05	0,09	X	
cons	0,00 – 0,02	0,04	X	
rati	0,00 – 0,03	0,05	X	
TOTAL			9	1

Gráfico 57



Infracciones y concordancias de 4 n-gramas en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Tabla 57

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
which	0,02 – 0,05	0,13	X	
ction	0,03 – 0,03	0,05	X	
ition	0,00 – 0,03	0,09	X	
ratio	0,00 – 0,02	0,04	X	
state	0,01 – 0,01	0,03	X	
latio	0,00 – 0,01	0,06	X	
ments	0,00 – 0,02	0,05	X	
ution	0,00 – 0,00	0,02	X	
tatio	0,00 – 0,01	0,03	X	
gener	0,00 – 0,00	0,02	X	
TOTAL			10	0

Gráfico 58



Infracciones y concordancias de 5 n-gramas en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Tabla 58

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O6	Infracciones	Concordancias
and	2,34 – 2,88	1,90	X	
with	0,56 – 0,77	0,76		X
be	0,36 – 0,51	0,52	X	
by	0,19 – 0,46	0,64	X	
this	0,03 – 0,33	0,83	X	
are	0,09 – 0,45	0,88	X	
which	0,08 – 0,22	0,63	X	
they	0,21 – 0,42	0,85	X	
you	1,00 – 1,87	0,23	X	
has	0,04 – 0,19	0,19		X
TOTAL			8	2

Gráfico 59



Infracciones y concordancias de palabras en la obra *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 6 del Corpus B.

Gráfico 60

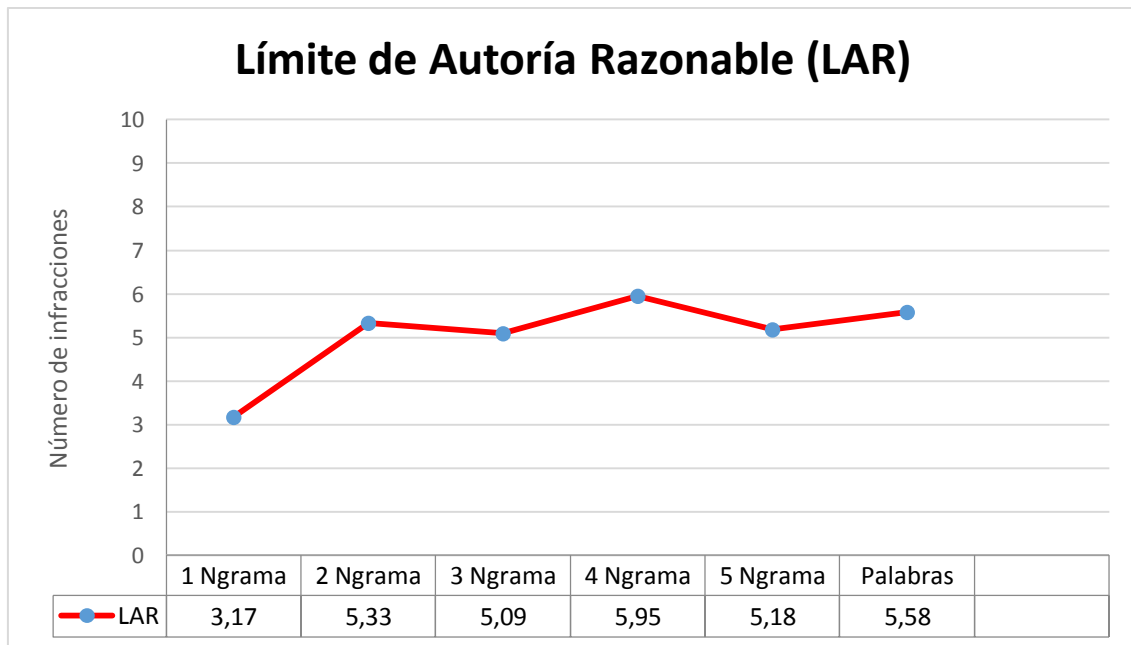


Gráfico 61

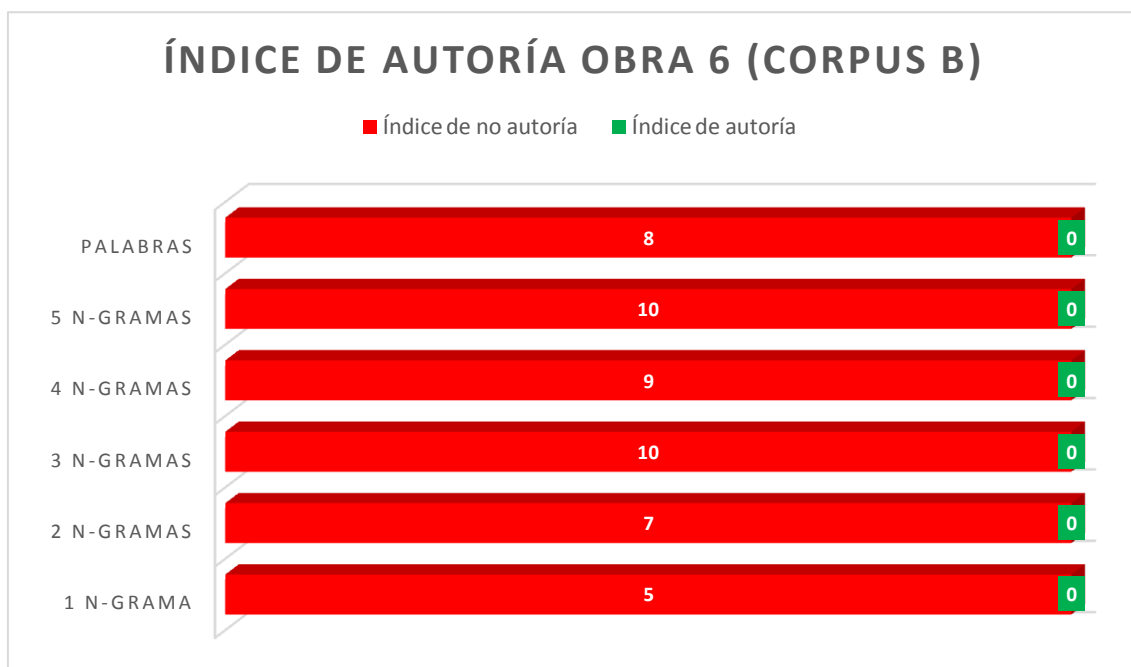


Gráfico 62

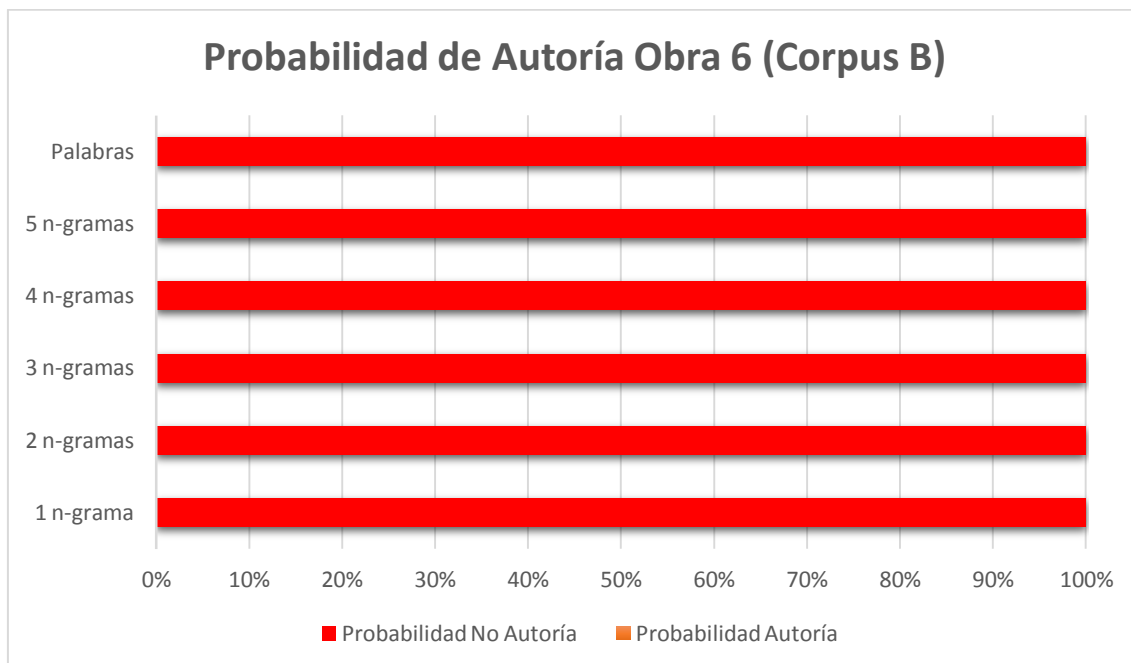
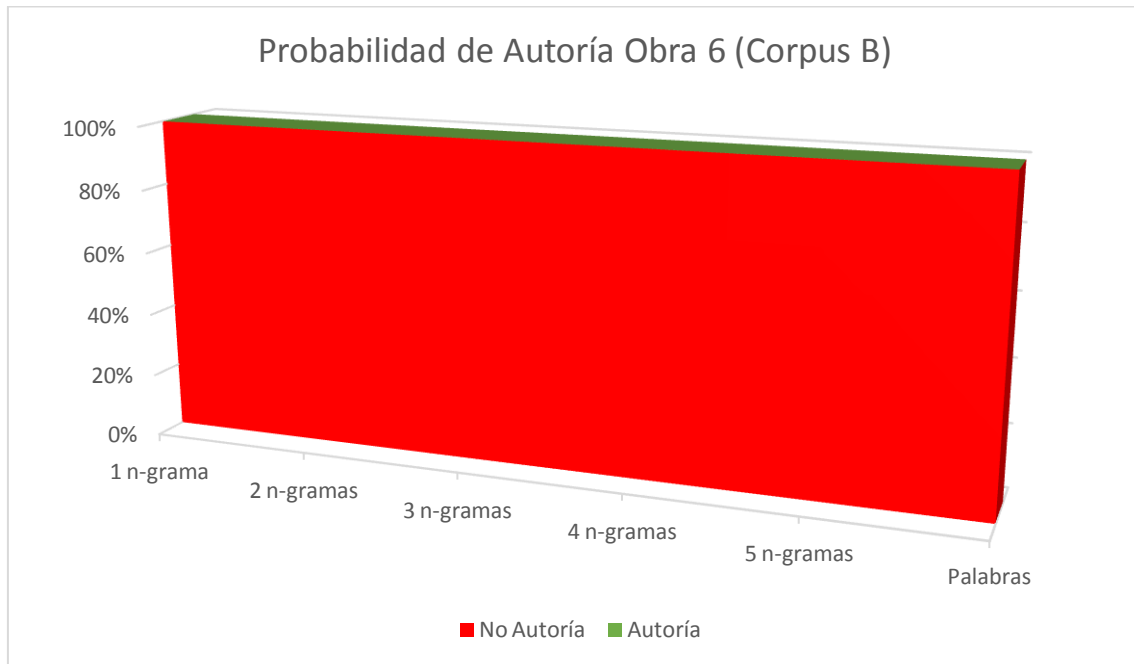


Gráfico 63



4.6.7. Análisis de la Obra 7 (Corpus B)

Título de la obra: *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about life, Death and the Universe*

Fecha de publicación: 1998

Tabla 59

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
“c”	1,93 – 2,35	2,45	X	
“s”	5,13 – 5,73	6,40	X	
“u”	2,76 – 3,18	2,61	X	
“p”	1,73 – 2,24	2,35	X	
“j”	0,22 – 0,34	0,19	X	
TOTAL			5	0

Gráfico 64

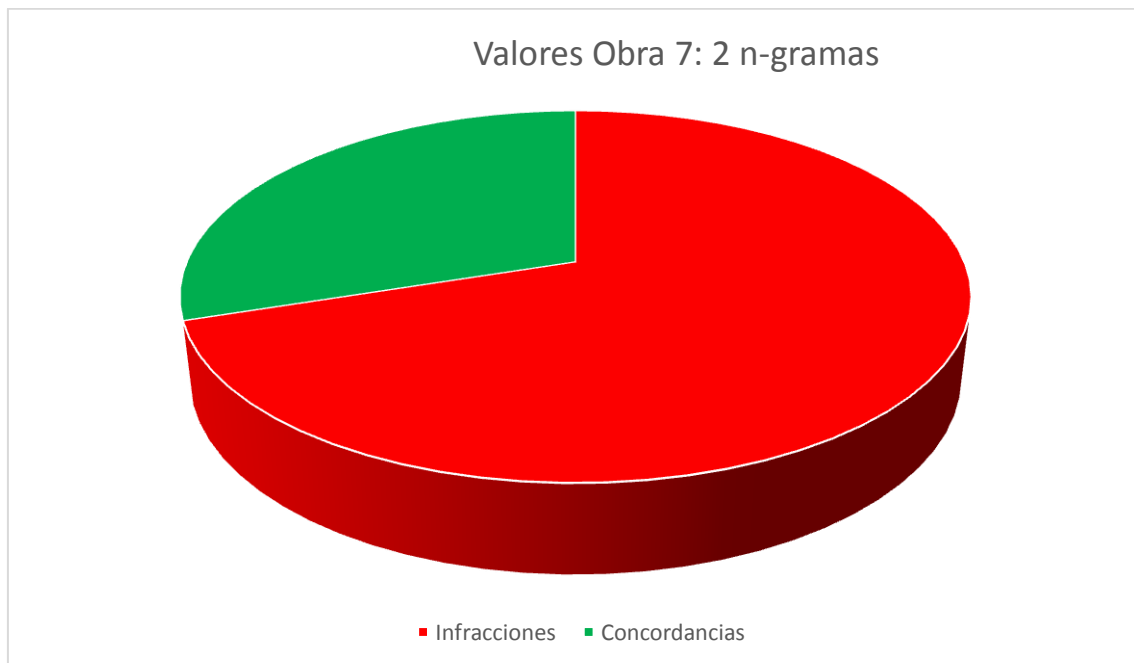


Infracciones y concordancias de 1 n-grama en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Tabla 60

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
er	1,31 – 1,43	1,38		X
nd	0,92 – 1,13	0,87	X	
of	0,48 – 0,75	0,88	X	
is	0,53 – 0,80	1,02	X	
as	0,79 – 0,88	0,76	X	
le	0,50 – 0,68	0,62		X
co	0,32 – 0,42	0,70	X	
ri	0,33 – 0,39	0,66	X	
ic	0,18 – 0,30	0,12	X	
ur	0,28 – 0,40	0,32		X
TOTAL			7	3

Gráfico 65



Infracciones y concordancias de 2 n-grama en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Tabla 61

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
and	0,66 – 0,78	0,53	X	
ion	0,13 – 0,34	0,34		X
con	0,10 – 0,11	0,14	X	
ons	0,05 – 0,08	0,16	X	
nce	0,11 – 0,14	0,20	X	
ive	0,05 – 0,11	0,17	X	
was	0,39 – 0,48	0,26	X	
ica	0,02 – 0,05	0,03		X
cti	0,03 – 0,06	0,10	X	
rat	0,03 – 0,06	0,05		X
TOTAL			7	3

Gráfico 66

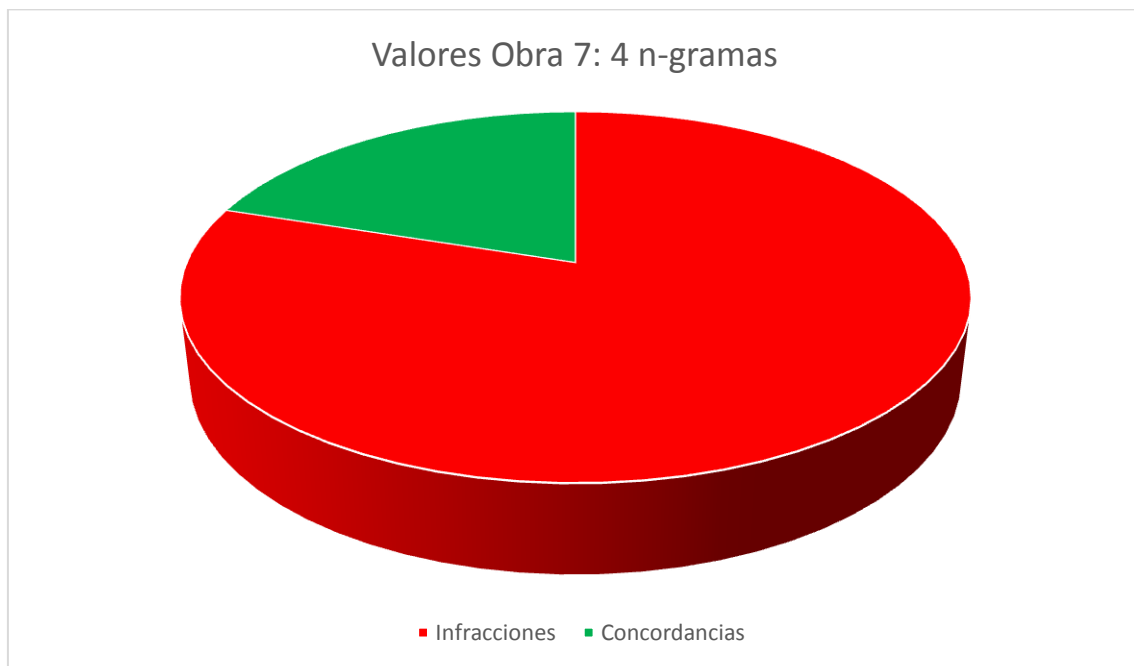


Infracciones y concordancias de 3 n-grama en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Tabla 62

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
ions	0,02 – 0,05	0,11	X	
hich	0,02 – 0,05	0,07	X	
ctio	0,02 – 0,04	0,05	X	
ight	0,08 – 0,14	0,07	X	
sion	0,03 – 0,03	0,04	X	
inte	0,02 – 0,05	0,08	X	
nter	0,00 – 0,06	0,05		X
heir	0,00 – 0,05	0,08	X	
cons	0,00 – 0,02	0,03	X	
рати	0,00 – 0,03	0,01		X
TOTAL			8	2

Gráfico 67

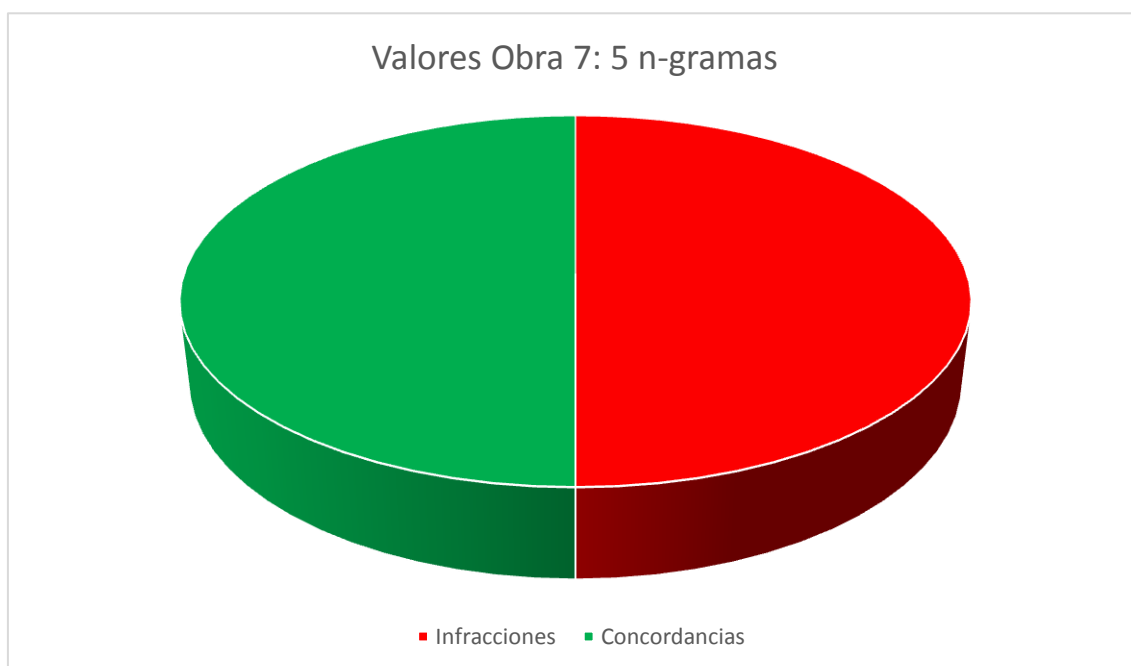


Infracciones y concordancias de 4 n-gramas en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Tabla 63

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
which	0,02 – 0,05	0,07	X	
ction	0,03 – 0,03	0,05	X	
ition	0,00 – 0,03	0,05	X	
ratio	0,00 – 0,02	0,01		X
state	0,01 – 0,01	0,02	X	
latio	0,00 – 0,01	0,01		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,04	X	
gener	0,00 – 0,00	0,00		X
TOTAL			5	5

Gráfico 68



Infracciones y concordancias de 5 n-grama en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Tabla 64

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O7	Infracciones	Concordancias
and	2,34 – 2,88	2,11	X	
with	0,56 – 0,77	0,60		X
be	0,36 – 0,51	0,72	X	
by	0,19 – 0,46	0,48	X	
this	0,03 – 0,33	0,33		X
are	0,09 – 0,45	0,50	X	
which	0,08 – 0,22	0,33	X	
they	0,21 – 0,42	0,57	X	
you	1,00 – 1,87	0,48	X	
has	0,04 – 0,19	0,34	X	
TOTAL			8	2

Gráfico 69



Infracciones y concordancias de palabras en la obra *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 7 del Corpus B.

Gráfico 70

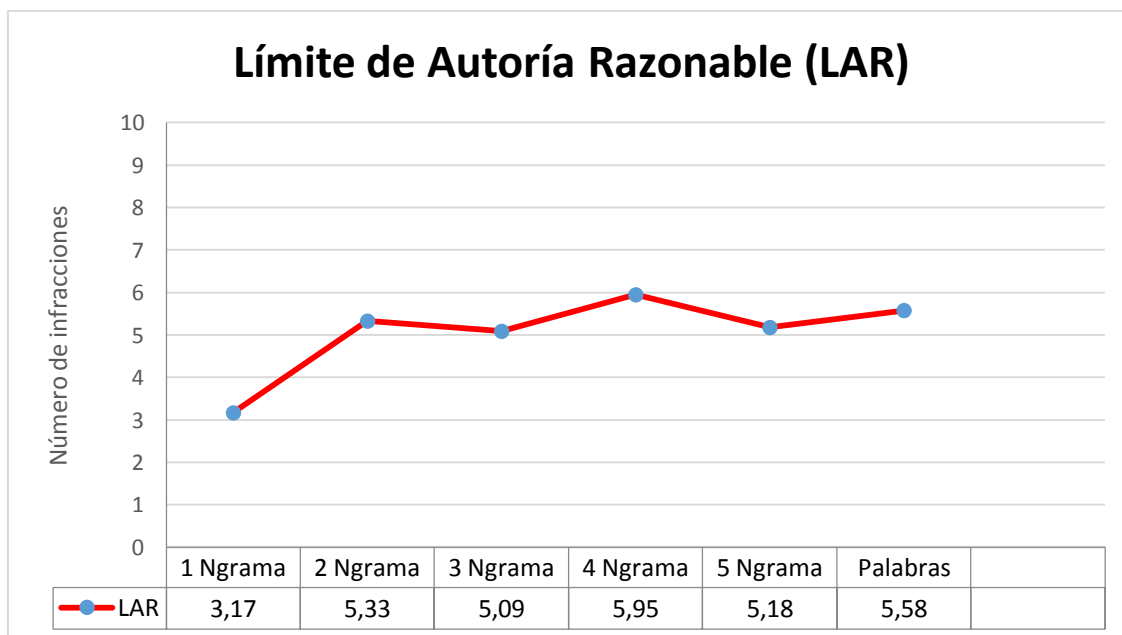


Gráfico 71

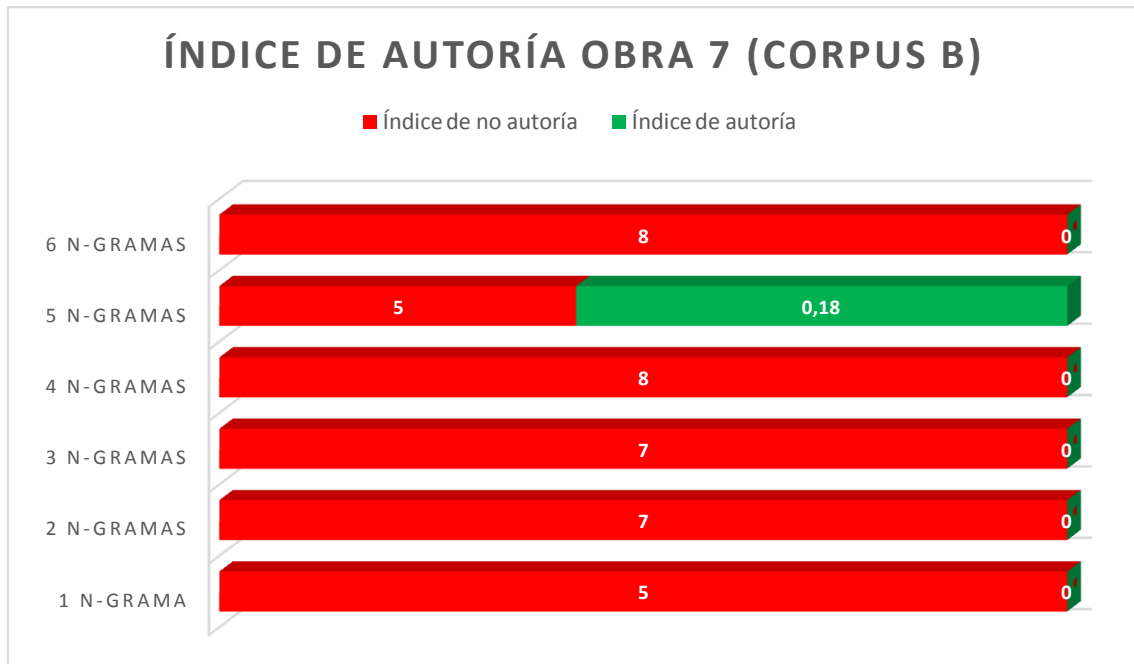


Gráfico 72

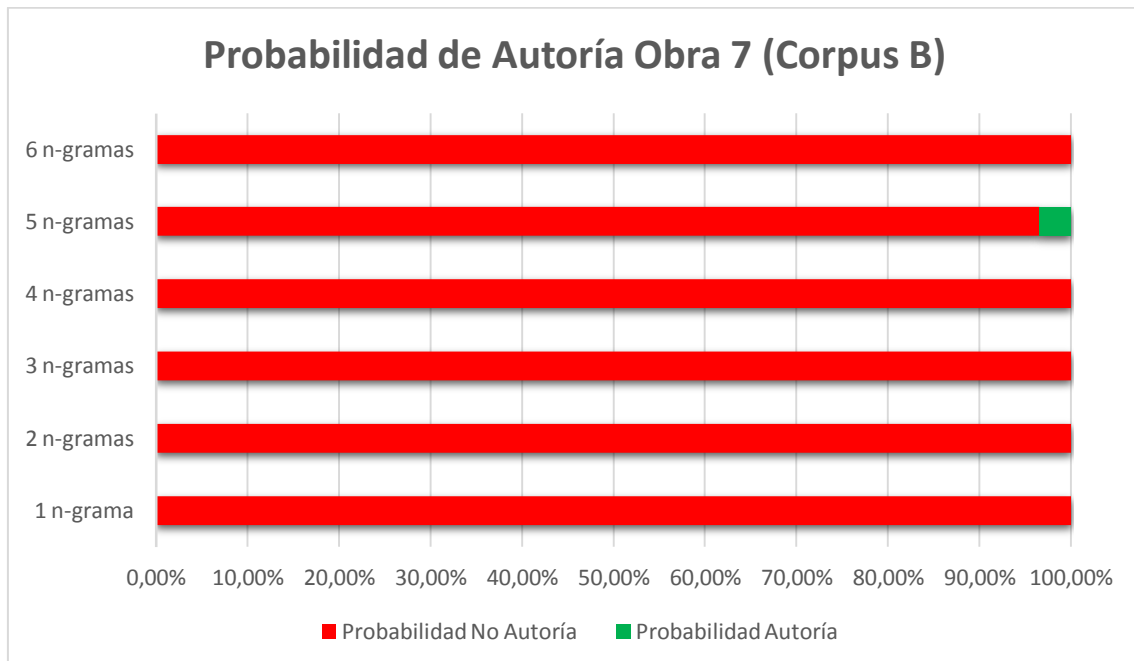
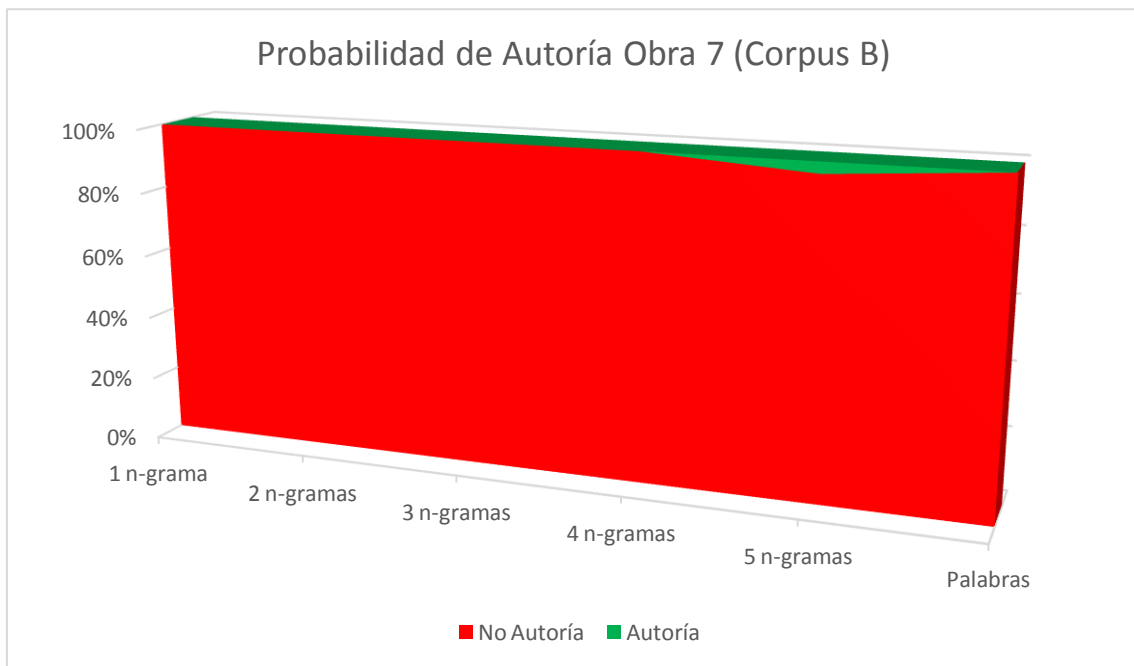


Gráfico 73



4.6.8. Análisis de la Obra 8 (Corpus B)

Título de la obra: *The Active Side of Infinity*

Fecha de publicación: 1998

Tabla 65

Categoría de marcador: 1 n-grama				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
“c”	1,93 – 2,35	2,32		X
“s”	5,13 – 5,73	5,67		X
“u”	2,76 – 3,18	2,81		X
“f”	1,73 – 2,24	2,18		X
“j”	0,22 – 0,34	0,24		X
TOTAL			0	5

Gráfico 74



Infracciones y concordancias de 1 n-grama en la obra *The Active Side of Infinity*

Tabla 66

Categoría de marcador: 2 n-gramas				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
er	1,31 – 1,43	1,48	X	
nd	0,92 – 1,13	0,88	X	
of	0,48 – 0,75	0,73		X
is	0,53 – 0,80	0,66		X
as	0,79 – 0,88	0,88		X
le	0,50 – 0,68	0,59		X
co	0,32 – 0,42	0,48	X	
ri	0,33 – 0,39	0,38		X
ic	0,18 – 0,30	0,31	X	
ur	0,28 – 0,40	0,19	X	
TOTAL			5	5

Gráfico 75

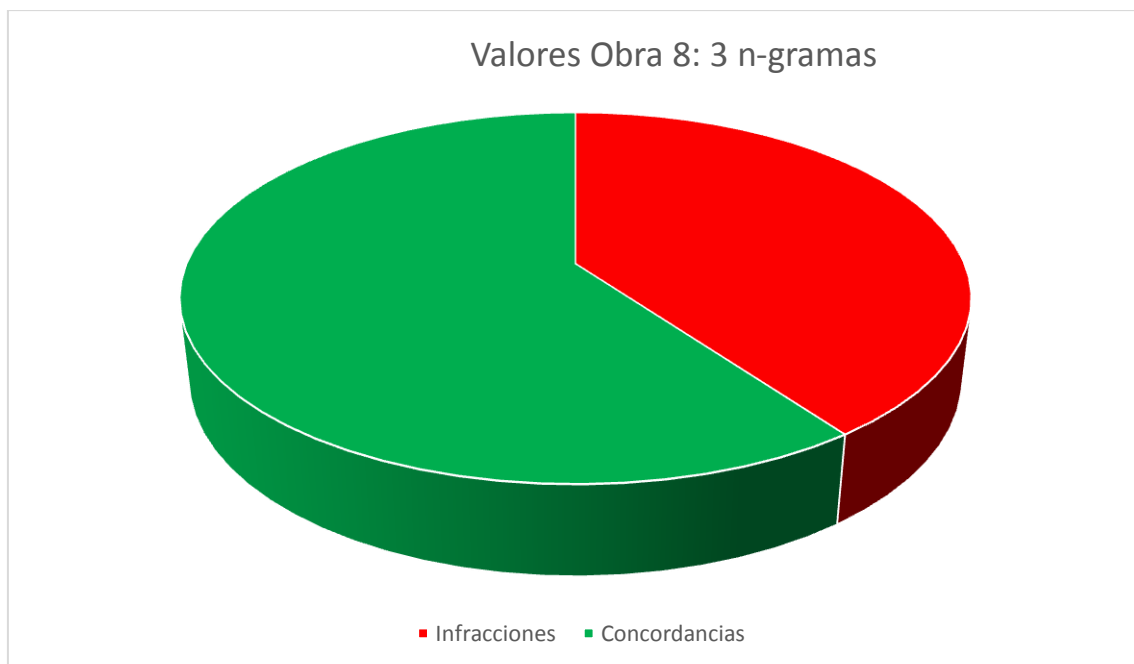


Infracciones y concordancias de 2 n-gramas en la obra *The Active Side of Infinity*

Tabla 67

Categoría de marcador: 3 n-gramas				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
and	0,66 – 0,78	0,53	X	
ion	0,13 – 0,34	0,28		X
con	0,10 – 0,11	0,13	X	
ons	0,05 – 0,08	0,09	X	
nce	0,11 – 0,14	0,14		X
ive	0,05 – 0,11	0,13	X	
was	0,39 – 0,48	0,43		X
ica	0,02 – 0,05	0,05		X
cti	0,03 – 0,06	0,06		X
rat	0,03 – 0,06	0,06		X
TOTAL			4	6

Gráfico 76

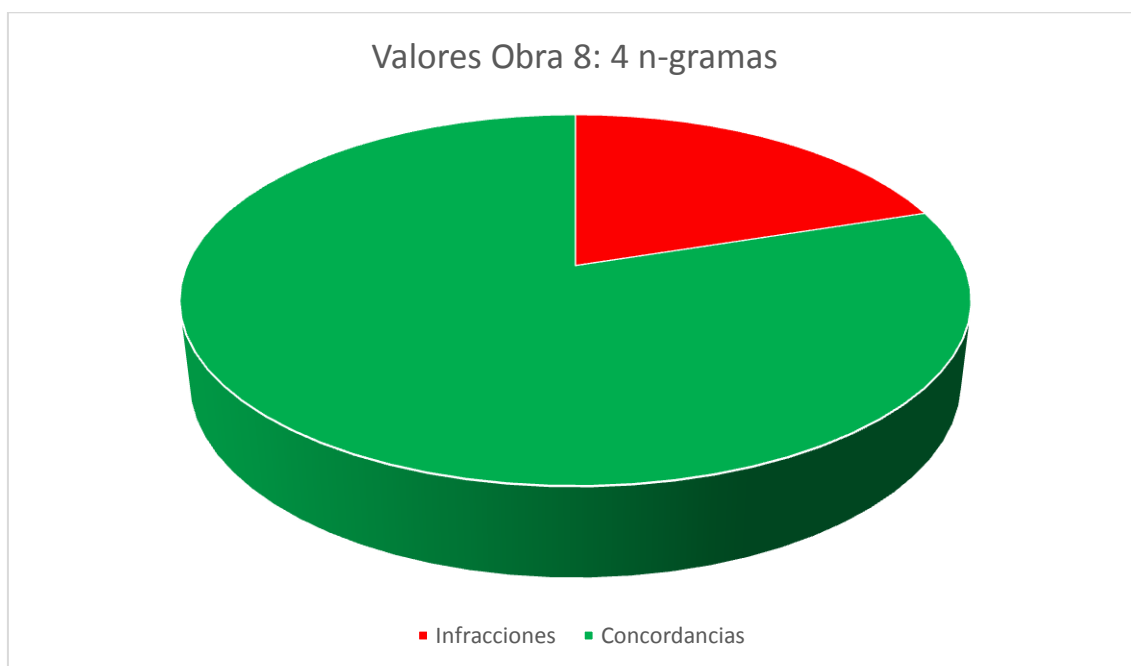


Infracciones y concordancias de 3 n-gramas en la obra *The Active Side of Infinity*

Tabla 68

Categoría de marcador: 4 n-gramas				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
ions	0,02 – 0,05	0,05		X
hich	0,02 – 0,05	0,51	X	
ctio	0,02 – 0,04	0,04		X
ight	0,08 – 0,14	0,08		X
sion	0,03 – 0,03	0,02	X	
inte	0,02 – 0,05	0,05		X
nter	0,00 – 0,06	0,04		X
heir	0,00 – 0,05	0,05		X
cons	0,00 – 0,02	0,02		X
рати	0,00 – 0,03	0,03		X
TOTAL			2	8

Gráfico 77

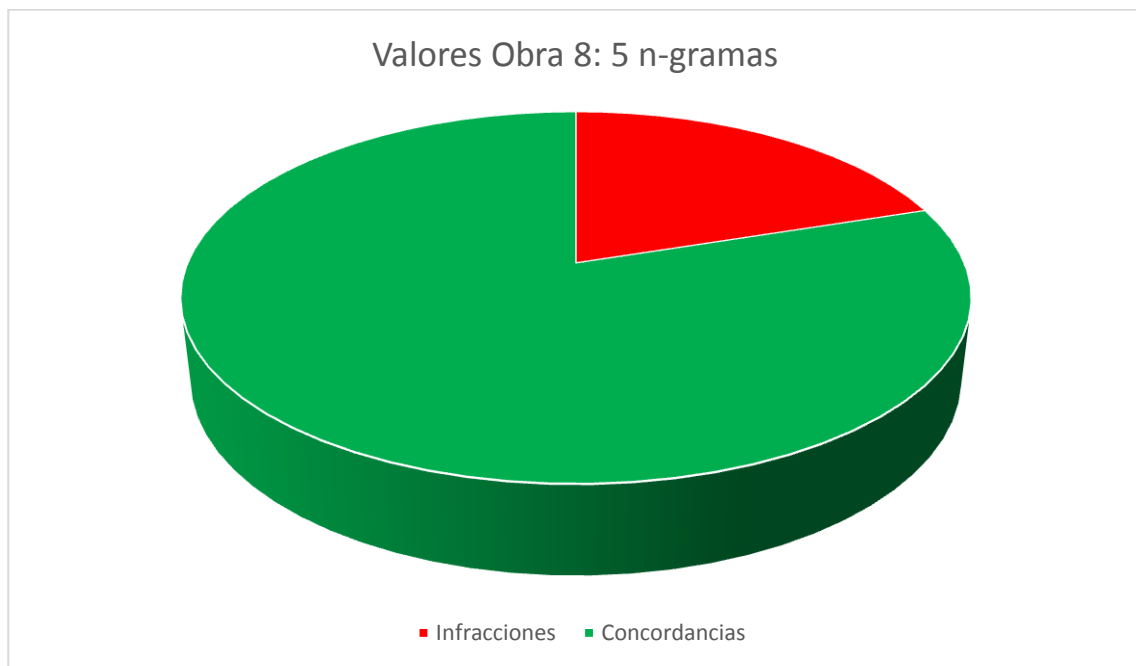


Infracciones y concordancias de 4 n-gramas en la obra *The Active Side of Infinity*

Tabla 69

Categoría de marcador: 5 n-gramas				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
which	0,02 – 0,05	0,05		X
Ction	0,03 – 0,03	0,04	X	
ition	0,00 – 0,03	0,05	X	
ratio	0,00 – 0,02	0,02		X
state	0,01 – 0,01	0,01		X
latio	0,00 – 0,01	0,01		X
ments	0,00 – 0,02	0,01		X
ution	0,00 – 0,00	0,00		X
tatio	0,00 – 0,01	0,01		X
gener	0,00 – 0,00	0,00		X
TOTAL			2	8

Gráfico 78



Infracciones y concordancias de 5 n-gramas en la obra *The Active Side of Infinity*

Tabla 70

Categoría de marcador: Palabras				
Marcador	Valor RdA	Valores O8	Infracciones	Concordancias
and	2,34 – 2,88	1,96	X	
with	0,56 – 0,77	0,70		X
be	0,36 – 0,51	0,40		X
by	0,19 – 0,46	0,35		X
this	0,03 – 0,33	0,37	X	
are	0,09 – 0,45	0,29		X
which	0,08 – 0,22	0,22		X
they	0,21 – 0,42	0,45	X	
you	1,00 – 1,87	1,19		X
has	0,04 – 0,19	0,10		X
TOTAL			3	7

Gráfico 79



Infracciones y concordancias de palabras en la obra *The Active Side of Infinity*

Contraste de autoría sobre la línea LAR

A continuación se contrastan los valores obtenidos en cada una de las categorías de marcadores con los de la línea LAR. En los gráficos se puede apreciar el porcentaje de probabilidad de que Carlos Castaneda sea el autor de la Obra 8 del Corpus B.

Gráfico 80

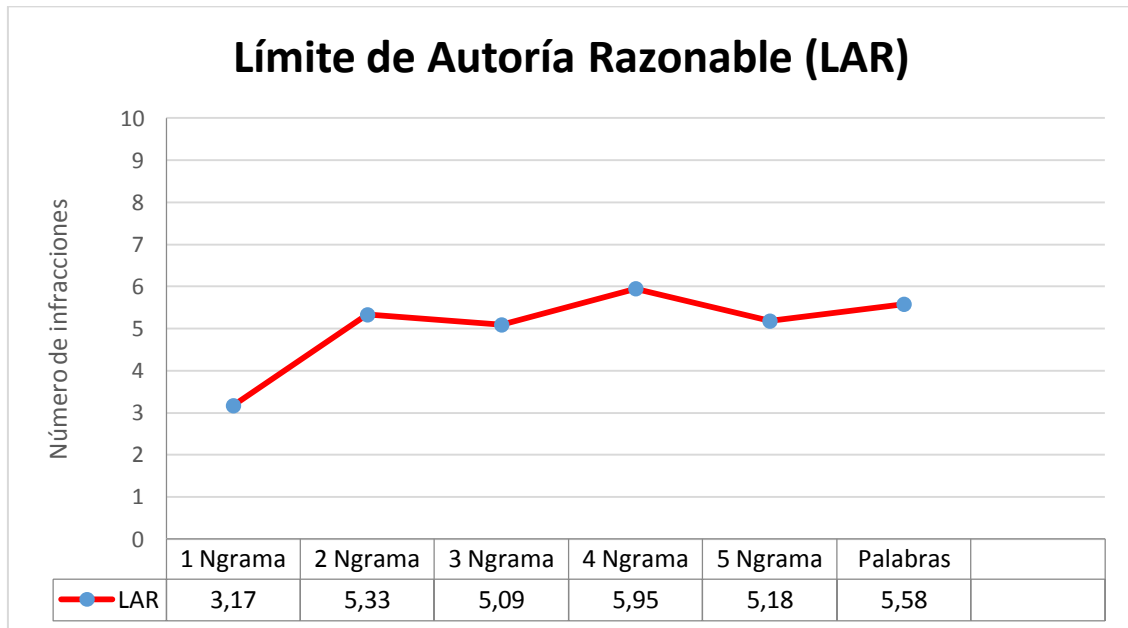


Gráfico 81

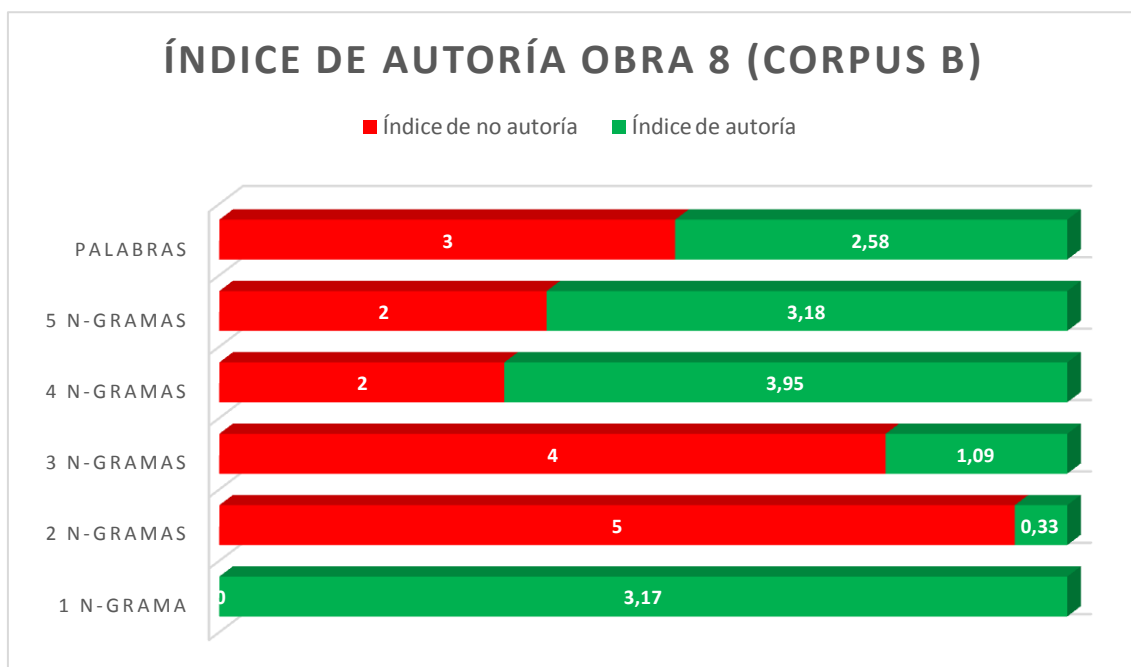


Gráfico 82

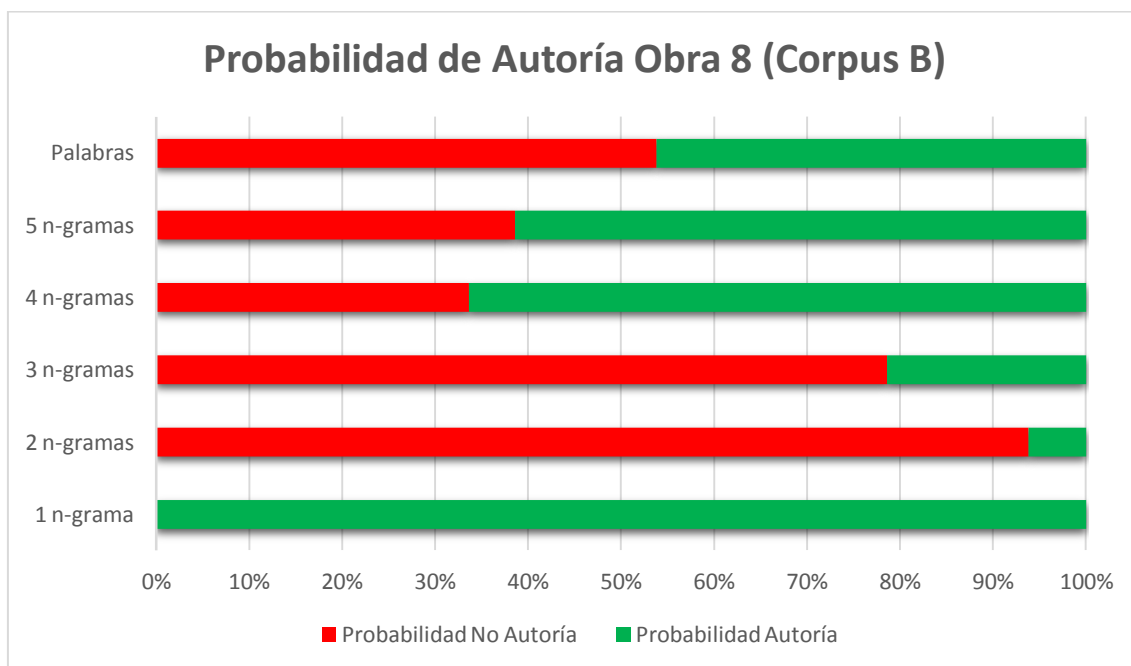
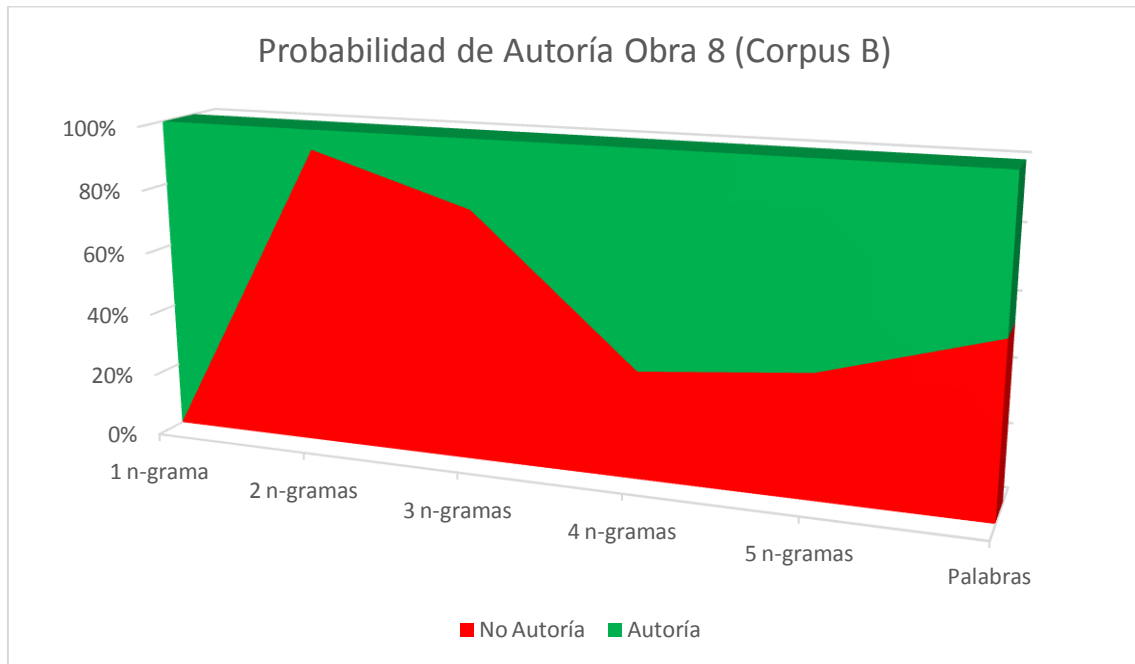


Gráfico 83

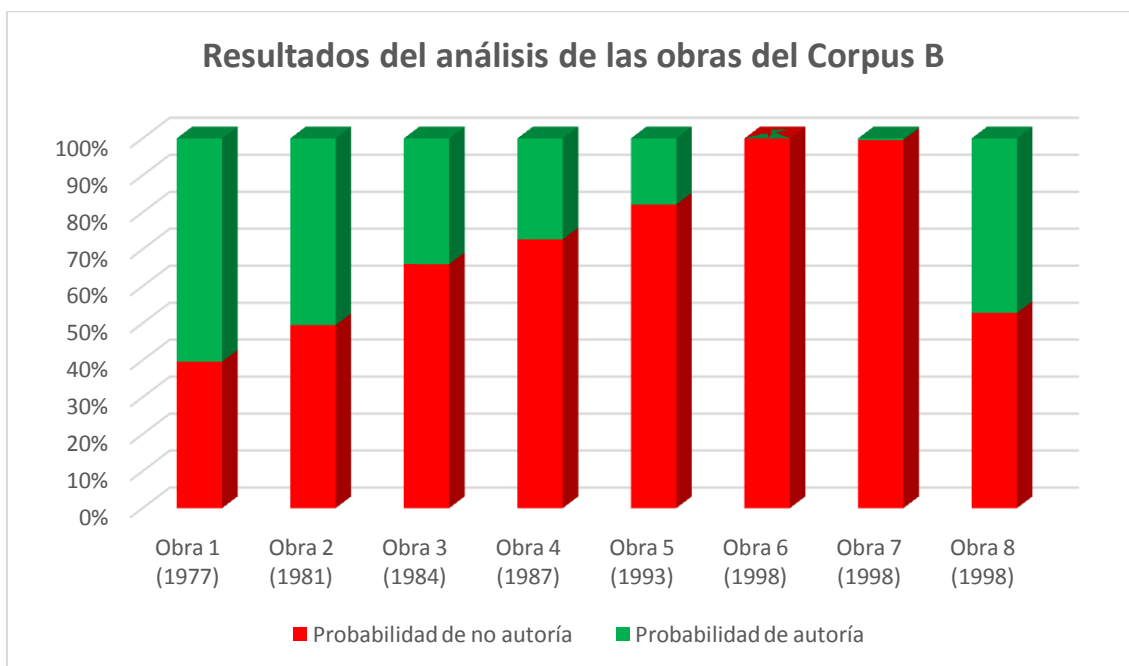


4.7. Fase 5. Determinación de porcentajes de autoría de las obras del Corpus B.

Como puede verse por los resultados que arrojan los análisis de cada una de las obras, los porcentajes de **probabilidad de no-autoría** varían de unas a otras y, en consecuencia, lo hacen igualmente solo que en proporción inversa los porcentajes de **probabilidad de autoría**.

La tabla que sigue muestra una visión del conjunto de las ocho obras y permite apreciar las diferencias resultantes una vez considerados el conjunto de los seis marcadores estilísticos utilizados en el análisis.

Gráfico 84



5. Análisis de los resultados y conclusiones generales

5.1. Análisis de los resultados: Sobre la autoría de Carlos Castaneda

Para llevar a cabo el estudio autorial de las obras de Castaneda hemos partido de anteriores estudios empíricos que muestran la enorme diversidad formal con que cada hablante expresa una misma idea. Esta diversidad constituye la esencia del idiolecto y permite, bajo ciertas condiciones, identificar al autor de un texto determinado.

En cualquier caso, la tarea de atribución de autoría se apoya en la hipótesis de que dos elementos que comparten un mismo origen, poseen también ciertas características que revelan este origen común. Trasladada esta premisa al terreno lingüístico equivale a decir que dos textos escritos por un mismo autor portan ciertos rasgos o elementos comunes que permiten reconocerlos como “hijos de un mismo padre”.

Por esta razón el primer paso de quien se propone atribuir un texto a un cierto autor consistirá en encontrar qué elementos se hallan presentes de forma regular en todos sus escritos. Esta será la primera condición para atribuir un texto dubitado a un autor candidato: que los rasgos de identidad que elijamos aparezcan siempre y en proporciones similares en los textos de dicho autor.

No todos los autores mostrarán la misma regularidad en su producción textual, por supuesto, y cuando más “irregulares” sean en su estilo de

escritura, más difícil le resultará al lingüista forense atribuirle un texto cualquiera.

En el caso que nos ocupa, además, no podíamos poner en práctica el método de la proximidad utilizado en aquellas situaciones en las que se cuenta con varios candidatos entre los cuales se tiene la certeza de que se encuentra el autor que buscamos. Se suelen resolver estos casos sin mayores dificultades atribuyendo el texto dubitado que estamos estudiando al escritor cuya producción textual guarde mayor similitud o proximidad. En este caso, y dado que tan solo contábamos con un autor candidato, no era posible aplicar este método, pues no hay forma de responder a la pregunta de cuánta similitud – o proximidad – entre los textos dubitados y los indubitados debíamos considerar suficiente para afirmar o negar la autoría del escritor analizado.

Decididos igualmente a evitar el método en el que se elabora un “panel de candidatos cerrado” de forma artificial por considerarlo poco fiable, optamos por elaborar un algoritmo específico para casos como el presente. Este algoritmo consta de tres etapas principales:

1. Determinar los Rasgos de Identidad del autor analizado (RdI)
2. Determinar el Rango de Autoría del dicho autor (RdA)
3. Establecer el Límite de Autoría Razonable (LAR)

En esencia, las dos primeras etapas determinan aquellos elementos que aparecen de forma regular en todos los textos escritos por un mismo autor y el grado de oscilación en la frecuencia de aparición que muestran dichos elementos. La tercera asume que la variación intra-autor es inherente a todo escritor y determina cuál es el grado esperable de dicha variación en el autor estudiado.

Las pruebas realizadas muestran que esta variación intra-autor es considerable, incluso en los autores más regulares, motivo por el cual incluso en textos producidos por un mismo autor encontraremos ejemplos de elementos que se “salen” del RdA. Cuando esto sucede decimos que se ha producido una “infracción” y para determinar cuántas infracciones son admisibles en el texto de un mismo autor llevamos a cabo un estudio estadístico con cuyos resultados elaboramos el LAR.

Así pues, una vez establecido este último punto aplicamos el algoritmo a las obras indubitadas de Carlos Castaneda, las que integran el Corpus A, para determinar los RdI y los RdA, y a las dubitadas, que integran el Corpus B, para comprobar si estas últimas superaban el número de “infracciones” establecido en el LAR.

Es decir, lo que buscábamos era determinar cuánto “se aleja” cada obra dubitada de las obras indubitadas y los resultados en cada una de las categorías de marcadores nos mostraron esta “distancia”.

Las obras que integran el Corpus B se clasificaron cronológicamente y siguiendo este mismo orden fueron analizadas. Ya hemos dicho que cada una de ellas se estudió por separado y de esta misma forma se muestran los resultados, que se desglosan por categorías de marcadores estilométricos. Al final del análisis de cada una de ellas se ofrecen varias tablas en las que es posible apreciar visualmente el grado de probabilidad / no probabilidad de autoría.

Por último, hemos mostrado un gráfico en el que es posible apreciar los resultados globales: el porcentaje total de autoría de cada obra – consideradas las seis categorías de marcadores – y el resultado de cada una de ellas en relación con todas las demás (gráfico 84).

De este modo, las cinco primeras obras muestran valores que, si bien difieren de los del Corpus A, se mantienen dentro de la tolerancia establecida por el LAR, aun cuando los porcentajes de probabilidad de no-autoría sean en algún caso mayores que los de autoría. Sin embargo, estos porcentajes arrojan unos valores absolutos en los casos de las obras 6 y 7. Tanto una como otra muestran que el porcentaje de probabilidad de que **no hayan sido escritas por Castaneda** es del 100%. En cambio, la última obra, la número 8, regresa a los valores aceptables, similares a los de las obras 2 y 3.

Una vez en este punto, cabe preguntarse cómo deben interpretarse estos resultados. Conviene recordar que estos porcentajes equivalen a la diferencia que muestra cada obra con respecto a aquellas que integran el Corpus indubitado. Los elementos estilométricos que hemos medido y que aparecen en los cuatro primeros libros de forma constante y regular, dejan de hacerlo, en mayor o menor medida, en los libros que integran la segunda etapa. Recordemos también que la atribución de autoría se basa precisamente en esto, en medir el grado de similitud entre distintos textos. Cuando las similitudes disminuyen más allá de ciertos límites, como es el caso de las obras 6 y 7, podemos asegurar con certeza que no han sido escritas por el autor candidato.

Ahora bien, al examinar en conjunto la obra de un mismo autor, como es este el caso, es preciso poner en contexto los datos obtenidos. Existen aquí ciertos factores extralingüísticos que condicionan necesariamente la forma en que deben ser contemplados los resultados del análisis estilométrico y el primero de ellos es la presencia de una pauta regular, de un patrón perceptible que regula los valores de no-autoría de cada obra dubitada del Corpus B.

Al situar dichos valores en un mismo gráfico y ordenarlos cronológicamente es posible apreciar que la “separación” de dichas obras con respecto a las del Corpus A aumenta a medida que se alejan en el tiempo.

Este distanciamiento paulatino pudiera ser debido a una evolución estilística, a un cambio progresivo de las técnicas de escritura, que se “alejan” de las utilizadas por Castaneda en el pasado.

Al igual que sucede con el idiolecto, el estudio diacrónico del estilo literario de un único individuo no ha recibido suficiente atención. Como ocurre con el habla, las diferencias entre los textos escritos por un mismo autor a lo largo del tiempo son tan irregulares e inconstantes que sin duda desaniman cualquier intento de encontrar un patrón en las mismas que permita sistematizarlas y elaborar una teoría que las explique. Pero la realidad es que en la práctica esta evolución estilística se produce y, por muy diversos motivos que sin duda varían de un escritor a otro, con el paso del tiempo el estilo cambia. En algunos casos este cambio puede ser apenas perceptible, pero en otros será tan acusado que pasados algunos años será difícil reconocer el estilo de un mismo escritor. Se trata de un factor que puede introducir cierta distorsión en los estudios de autoría y contra el que ya prevenía Koppel (2009) como vimos al principio de este trabajo.

Este podría ser el caso de Carlos Castaneda. Cada una de las cinco primeras obras se aleja un poco de la anterior hasta que finalmente la sexta y la séptima aparecen tan distantes, tan diferentes de acuerdo a nuestras mediciones, que no es de extrañar que parezcan haber sido escritas por alguien diferente.

Es cierto, también, que a partir de la obra 4 y hasta la obra 7 los porcentajes de no-autoría son elevados, sin embargo, es importante recordar que esos porcentajes representan la distancia entre dichas obras y el Corpus A. En cambio cuando comparamos cada una de ellas con la inmediatamente anterior las distancias se atenúan y en conjunto muestran un paulatino alejamiento del estilo inicial.

Tabla 71

Distancia entre las obras del corpus B							
Obra 1	Obra 2	Obra 3	Obra 4	Obra 5	Obra 6	Obra 7	Obra 8
39,6%	49,5%	66%	72,6%	82,11%	100%	99,55%	52,8%

Si comparamos cada obra con la inmediatamente anterior vemos que la distancia entre ambas es tan escasa que no cabe sino considerarlas obra del mismo autor.

Por tanto, el hecho de que la separación entre las siete primeras obras del Corpus B aumente de forma progresiva, indica que se ha producido una evolución estilística. Esta evolución explicaría, a nuestro parecer, los elevados porcentajes de probabilidad de no-autoría de las obras 6 y 7.

Sin embargo, es preciso considerar otros factores adicionales. El inicio de la segunda etapa de la producción literaria de Castaneda coincide en el tiempo con su alejamiento de la vida pública; entre la publicación de la última obra del Corpus A, *Tales of Power* (1974) y la primera del Corpus B, *The Second Ring of Power* (1977), transcurren solo 3 años, pero tanto el contenido de los libros como las circunstancias personales del autor cambian drásticamente. Cuando, al final de *Tales of Power*, Castaneda salta desde un risco de las montañas del norte de México a un abismo de oscuridad, concluyendo así su aprendizaje y despidiéndose de su mentor, don Juan, el autor salta metafóricamente a un precipicio literario obligándose a sí mismo a cambiar los esquemas narrativos empleados hasta entonces. Si las cuatro primeras obras estaban formadas casi exclusivamente por los diálogos maestro-alumno y los vagabundeos de ambos por los páramos de la región de Sonora, a partir de entonces el autor se ve forzado a introducir nuevos personajes, nuevos escenarios y nuevos, y cada vez más extraños, conflictos que mantengan la tensión narrativa.

En lo personal, además, Castaneda se retira a su refugio místico en compañía de sus acólitas, corta los vínculos con el mundo exterior y lo cierto es que nadie sabe muy bien en qué ocupa su tiempo durante los años que siguen. Para la persona que Castaneda pretende ser supone un paso importante, decisivo, pues a partir de ese momento debe representar durante las veinticuatro horas del día el papel que para sí ha creado. ¿Llega en algún momento a creerse su propia historia? No hay forma de saberlo, pero está claro que se ve obligado a introducir cambios y ajustes en el mundo de ficción – supuestamente verídico – que ha elaborado. Mediante una pirueta narrativa trae de vuelta a don Juan a sus libros⁷¹, gracias a la estratagema literaria del acceso a una parte de su subconsciente que guarda recuerdos que han permanecido dormidos hasta ese momento, como se mencionaba en el capítulo 3.4.

Todo ello ha de haber influenciado, sin duda, al Castaneda escritor, reflejándose en el estilo de los libros que produce a partir de entonces. No es de extrañar, pues, que entre *Tales of Power* y *The Second Ring of Power* exista una diferencia estilística que hace alcanzar a esta segunda casi un 40% de probabilidad de no-autoría.

Pero existen, además, otros factores contextuales que parecen relevantes, como por ejemplo el lapso temporal discordante que separa los primeros cinco libros del Corpus B y el que separa los tres últimos. Tras su retiro, Castaneda publica sus siguientes cinco libros con un intervalo que oscila entre 3 y 6 años entre cada uno de ellos. Su reaparición coincide con la publicación del quinto libro, en 1993, que el escritor aprovecha para promocionar con charlas públicas. Los años siguientes multiplica sus conferencias, seminarios y entrevistas en diversos medios de comunicación, aunque hacia finales de 1995 su salud comienza a deteriorarse.

⁷¹ En su siguiente libro, *The Second Ring of Power*, el lector descubre que Don Juan ha saltado también al abismo, después de que lo hiciera Carlos Castaneda, pero el maestro, a diferencia del discípulo, se traslada a otro plano de la existencia, desapareciendo de este mundo.

Los tres últimos libros se ponen a la venta el mismo año de su fallecimiento, 1998, en lo que parece un intento de rentabilizar al máximo la figura del gurú espiritual en que se ha convertido Carlos Castaneda. Habían pasado cinco años desde la aparición del anterior, el quinto, *The Art of Dreaming* (1993) y la disonancia entre el tiempo que hasta entonces se tomaba Castaneda entre un libro y otro y este repentino incremento en la productividad, es un motivo más que hace sospechar la intervención de personas ajenas a Castaneda en la elaboración de estos volúmenes.

Hay todavía otro factor más que ha de tenerse en cuenta: el regreso de la obra 8, la última del Corpus B, a valores intermedios. La repentina vuelta a un valor que se sitúa entre las obras 2 y 3, distanciándose bruscamente de las más recientes, es un indicio claro de que hay algo irregular en su elaboración. Si el estilo de Castaneda fue experimentando un cambio paulatino y consistente a medida que pasaban los años, alejándose cada vez más del Corpus A, ¿cómo explicar este nuevo cambio, esta vez sí radical que retrocede y vuelve a acercarse a las obras indubitadas?

La obra 8, *The Active Side of Infinity*, comienza con estas palabras: “This book is a collection of the memorable events in my life.” En sí mismas no son especialmente reveladoras, pues en buena parte de los libros que componen este Corpus B el autor rememora acontecimientos pasados y también en el octavo Castaneda narra una vez más historias personales y distintos sucesos más o menos significativos. Sin embargo, como ya hemos dicho, este libro apareció pocos meses después de su fallecimiento, casi a la vez que el séptimo, *The Wheel of Time*, y menos de un año desde la publicación del sexto, *Magical Passes*. Por todo ello, no parece descabellado suponer que alguien de su entorno cercano hizo uso de sus notas, apuntes o borradores escritos tiempo atrás, pero nunca publicados, para componer un último libro bajo el nombre de Carlos Castaneda. Estilométricamente el texto se aproxima a los de la década anterior, años en los que, a juzgar por el contenido, el autor revisó una y otra

vez lo narrado en los cuatro primeros libros y recompuso nuevas experiencias y andanzas con su maestro, intercalándolas de forma cada vez más inverosímil entre las “originales” del Corpus A. Esto coincide con el contenido de la obra 8 y explicaría la similitud entre esta pieza y las obras 2, *The Eagle's Gift* (1981), y 3, *The Fire From Within* (1984).

Ya vemos, pues, que a pesar de que los resultados de las mediciones y contrastes de los elementos característicos de cada una de las obras arrojan unos resultados concretos, al evaluar el conjunto de las circunstancias que rodean su elaboración la perspectiva cambia considerablemente. Por una parte, tomados individualmente los resultados de cada una de las obras está claro cuáles pueden ser atribuidas a Castaneda, cuáles suscitan dudas y de cuáles se puede afirmar con seguridad que han sido escritas por otro autor.

Por otro lado, al examinarlas todas ellas en conjunto y al considerar el contexto en que se elaboran, surgen explicaciones plausibles para esos resultados, que proporcionan una visión muy distinta. Se trata, por tanto, de un caso muy especial que requiere sopesar las pruebas a favor y en contra antes de emitir un juicio y surgen tres hipótesis, cada una de ellas con una interpretación diferente de los mismos hechos.

Análisis de los resultados: Hipótesis 1.

En la segunda etapa literaria de Carlos Castaneda, debido muy probablemente a sus circunstancias personales y a la necesidad de iniciar un nuevo ciclo narrativo, el estilo de escritura comienza a cambiar y sigue haciéndolo paulatinamente a lo largo de los años siguientes. Por esta razón, el análisis estilométrico de las obras que componen el Corpus B muestra unos valores que se alejan del Corpus A, a medida que pasa el tiempo.

Esta distancia estilística incrementa progresivamente los valores de probabilidad de no-autoría hasta alcanzar el valor máximo en las obras 6 y 7. Por lo tanto no cabe considerar a ninguna de las siete primeras obras ajena a la pluma del escritor analizado.

En cuanto a la última obra, la 8, se trataría de una obra escrita tiempo atrás, pero nunca publicada. Esto explicaría su regreso a unos valores a medio camino entre los de las obras 2 y 3.

Tabla 72

Hipótesis 1		
OBRAS	CONCLUSIÓN	MOTIVOS
1, 2, 3,4 ,5, 6 y 7	Atribuibles a Carlos Castaneda	Los elevados porcentajes de no-autoría de alguna de ellas, en especial de la 6 y la 7, son debidos a la evolución estilística progresiva , como se puede apreciar en el incremento paulatino de dichos valores.
8	Atribuible a Carlos Castaneda	La diferencia brusca de porcentajes de no-autoría con respecto a obras inmediatamente anteriores sugiere que se trata de una obra antigua, escrita por el propio Castaneda, si bien nunca publicada y recuperada de forma póstuma.

Análisis de los resultados: Hipótesis 2

De nuevo la evolución estilística del escritor explicaría los valores crecientes de probabilidad de no-autoría. El estilo de Castaneda cambia con el tiempo culminando en las obras 6 y 7, cuyos altos porcentajes deben atribuirse a este cambio estilístico y no a la intervención de otro autor.

Por el contrario, la obra 8 retrocede de manera inesperada hasta los valores de las obras escritas más de diez años atrás. Si la obra fue escrita en esa época, no parece haber ningún motivo lógico por el que hubiese de permanecer inédita durante tanto tiempo. Cabe considerarla por tanto, y a pesar de sus moderados valores de probabilidad de no-autoría, como atribuible a otro escritor distinto de Castaneda.

Tabla 73

Hipótesis 2		
OBRAS	CONCLUSIÓN	MOTIVOS
1, 2, 3, 4, 5, 6 y 7	Atribuibles a Carlos Castaneda	Los elevados porcentajes de no-autoría de alguna de ellas, en especial de la 6 y la 7, son debidos a la evolución estilística progresiva , como se puede apreciar en el incremento paulatino de dichos valores.
8	No atribuible a Carlos Castaneda	A pesar del moderado porcentaje de no-autoría, la diferencia repentina con respecto a las obras inmediatamente anteriores sugiere la intervención de un segundo autor

Análisis de los resultados: Hipótesis 3

De acuerdo a esta hipótesis, la evolución estilística justificaría el incremento de probabilidad de no-autoría de las cinco primeras obras del corpus B. Sin embargo, el hecho de que las tres últimas aparezcan en un plazo de tiempo muy breve, en el mismo año de su fallecimiento, hace sospechar la intervención de un segundo autor en la elaboración de la 6 y de la 7.

La 8, por su parte, muestra un brusco descenso con respecto incluso de la obra 5, la última considerada atribuible a Castaneda, según esta hipótesis, y

se explicaría si alguien hubiese tomado notas y borradores antiguos no utilizados para componer un último libro, publicado algunos meses después de su fallecimiento.

Tabla 74

Hipótesis 3		
OBRAS	CONCLUSIÓN	MOTIVOS
1, 2, 3,4 y 5	Atribuibles a Carlos Castaneda	Los elevados porcentajes de no-autoría de alguna de ellas, como la 4 y la 5, son debidos a la evolución estilística progresiva , como se puede apreciar en el incremento paulatino de dichos valores
6 y 7	No atribuibles a Carlos Castaneda	Porcentajes del 100% de no-autoría sumados al inusualmente breve lapso de tiempo en que se publican.
8	Atribuible parcialmente a Carlos Castaneda	A pesar del moderado porcentaje de no-autoría, la diferencia repentina con respecto a las obras inmediatamente anteriores sugiere que un segundo autor habría utilizado borradores y notas antiguas del propio Castaneda para “componer” este último libro.

Como se ve, todas las hipótesis asumen la evolución estilística como el factor responsable del aumento de la probabilidad de no-autoría. Las diferencias aparecen en el momento de evaluar los resultados de las obras 6, 7 y 8, pues se ven condicionadas por los elementos contextuales descritos más arriba.

Por nuestra parte, consideramos que la tercera es la hipótesis más probable y ello debido al fuerte peso que, a nuestro parecer, tiene el hecho de que estas tres últimas obras apareciesen en tan breve espacio de tiempo. El mismo año en que Castaneda fallece, se publican *Magical Passes*, *The Wheel*

of Time y *The Active Side of Infinity*. La primera apenas dos meses antes de su fallecimiento y las otras dos posteriormente. El lapso de tiempo entre la aparición de cada uno de los anteriores cinco libros oscila entre tres y seis años; en este caso, en cambio, tras cinco años de ausencia literaria aparecen tres obras prácticamente a la vez y coincidiendo con su fallecimiento. Sin duda hay en este movimiento una calculada estrategia editorial. Agotada la posibilidad de seguir en lo sucesivo explotando comercialmente al grupo de lectores incondicionales de Castaneda, alguien decide rentabilizar una última vez – tres en realidad – el nombre del escritor.

Magical Passes es una especie de manual técnico de *gestos* o *movimientos mágicos*, tal y como su título indica. En él se instruye al lector en movimientos, posturas y gestos que tendrían la finalidad de “restaurar la energía del individuo”, de acuerdo a las explicaciones del propio autor. A esas alturas, Castaneda había desarrollado un sistema de movimientos y ejercicios físicos que él y sus discípulas transmitían habitualmente en cursos y seminarios y al que habían dado el nombre de “*tensegrity*”. Al tratarse, por tanto, del equivalente a un libro técnico podría haber sido redactado sin mucha dificultad por alguien de su entorno que colaborase habitualmente en la impartición de dichos seminarios.

The Wheel of Time, en cambio, contiene citas y pasajes tomados de libros anteriores seguidos de los correspondientes comentarios del autor. De nuevo, no parece difícil que alguien de su entorno, acostumbrado a “interpretar” el contenido de sus libros y a transmitirlo en los mencionados seminarios, realizase esta labor de exégesis y firmase con el nombre de Castaneda.

Por último, en *The Active Side of Infinity* el autor, en compañía de don Juan, rememora acontecimientos pasados, hechos supuestamente trascendentes de su vida. Se trata de un esquema que se ajusta a lo que se

había convertido en habitual en la segunda etapa literaria de su vida: volver atrás en el tiempo y “recuperar” diálogos inéditos entre Castaneda y su maestro. Además, el escritor siempre había sido aficionado a incluir en sus libros anécdotas propias, ninguna de las cuales pudo verificarse nunca. Incluso su última entrevista, concedida apenas seis meses antes de su desaparición, está llena de pequeñas historias personales (Brenan, 1997). Así que tampoco aquí resulta descabellado suponer que se trate de escritos del propio Castaneda, pruebas y borradores con escenas desarrolladas tiempo atrás, pero que por una u otra razón no llegó a incluir en sus libros anteriores. Sin embargo, si como indican los análisis estilométricos, el libro fue escrito en una época previa, si estaba completo, ¿por qué no llegó nunca a publicarse? Parece más razonable pensar que alguien de su entorno cercano hubiese recogido esos textos dispersos de entre sus papeles personales y les hubiese dado forma, elaborando el que sería el último libro del maestro, quien ya no escribiría más. La autoría del mismo, por tanto, debería atribuirse solo parcialmente a Castaneda.

5.2. Conclusiones generales: Dos factores para su consideración en los estudios de atribución de autoría.

Como puede verse, el estudio de autoría sobre las obras del escritor Carlos Castaneda aparece, en su etapa final, lleno de incertidumbres y ello a pesar de que los resultados de nuestros análisis estilométricos proporcionan cifras y datos muy precisos. Sin embargo, y aunque el resultado final de este trabajo concreto resulte hasta cierto punto frustrante, las conclusiones que de él se pueden extraer nos parecen de gran valor y aplicables a la generalidad de estudios de autoría. Estas conclusiones pueden resumirse en dos puntos específicos:

1. Los estilos de escritura varían con el paso del tiempo. La misma obviedad de esta afirmación puede llevar a olvidarla a quien intenta determinar el autor de una pieza de texto, pero lo cierto es que debe ser tomada muy en cuenta. Como decíamos más arriba, la dificultad para sistematizar esta variación, esta evolución estilística, no hace sino complicar la tarea del investigador. El estilo no varía en todos los escritores en la misma proporción, ni al mismo ritmo, por lo que cada caso debe ser examinado individualmente. Pero el hecho es que en ciertos autores la deriva estilística puede llegar, con el paso de los años, al punto de hacer pensar que dos textos procedentes del mismo escritor fueron generados por personas distintas y el caso de Castaneda es buen ejemplo de ello. Así pues, al plantearnos un estudio de este tipo habría que determinar si las muestras, dubitadas e indubitadas, proceden de fechas cercanas y de no ser así, considerar la posible evolución estilística como un factor que puede condicionar la certeza del veredicto final.

2. El contexto juega un papel relevante en la interpretación de los resultados. Como se ha visto en esta ocasión, los factores extralingüísticos han de ser tomados en cuenta a la hora de sopesar el resultado del análisis. Las circunstancias en las que los textos fueron elaborados pueden proporcionar explicaciones que justifiquen de manera razonable cifras que en principio parecen apuntar con claridad en una dirección determinada. Por poner una analogía, los análisis de sangre de un deportista pueden mostrar la presencia de sustancias prohibidas, pero es posible que este deportista haya estado tomando una medicación que contenga dichas sustancias, por lo que las sanciones reglamentarias deberán ser reconsideradas. De igual forma, resultados que sugieren la existencia de un segundo autor pueden deberse a factores ajenos a los textos analizados. Y a la inversa, por mucho que una obra concreta se sitúe dentro del rango de autoría, si existe una distancia acusada con los valores de otra obra cercana en el tiempo del mismo autor, esto es indicio de alguna irregularidad que merece la pena investigar.

Todo esto no hace sino sugerir que la rama de la lingüística forense que pretende determinar la autoría de textos escritos a partir de su análisis estilométrico debe asumir un cierto grado de incertidumbre y relatividad en sus resultados. En ocasiones, las conclusiones a las que se llega tras el estudio de los textos dubitados serán tan precisos como los propios resultados estadísticos, pero asumir que siempre será así conducirá, antes o después, al error. Los programas de cálculo y de análisis de datos cada vez más perfeccionados contribuirán a incrementar esa precisión, si bien las circunstancias que rodean la elaboración de los textos analizados siempre deberán ser tenidas en cuenta y no parece probable que terminen siendo susceptibles de su conversión a elementos numéricos. Esto significa que la tarea del investigador humano continuará siendo necesaria en el futuro y la experiencia acumulada en trabajos anteriores facilitará el trabajo final de interpretación.

En cuanto al valor literario de la obra de Castaneda, nos parece que este debe ser apreciado de forma independiente prescindiendo de cuestiones como el origen de la misma o su veracidad. Una vez asumido su carácter de “ficción” y olvidada toda pretensión de documento antropológico. Revisiones más calmadas sin duda pondrán de manifiesto que contiene elementos de una elevada calidad poética, capaces de emocionar al lector y hacerle soñar con la posibilidad de mundos cuya existencia está más allá de la razón. Su caso bien podría seguir una evolución similar a la que experimentaron en su momento la obra de James McPherson y sus ciclos ossianicos, casi tres siglos antes, cuando la constatación de su impostura arqueológica, dejó paso paulatinamente a su aprecio como obra poética e incluso al reconocimiento de su aportación a la cultura gaélica. El don Juan de Castaneda, al igual que Ossian, pertenece ya al imaginario colectivo y es allí donde reside su verdadero valor.

6. Bibliografía

6.1. Fuentes Primarias Consultadas:

Castaneda, Carlos ([1968] 1998). *The Teachings of Don Juan: A Yaqui Way of Knowledge*, 30th Anniversary Edition. New York: Washington Square Press.

Castaneda, Carlos (1971). *A Separate Reality*. New York: Pocket Books.

Castaneda, Carlos (1972). *Journey to Ixtlan*. New York: Simon & Schuster.

Castaneda, Carlos (1974). *Tales of Power*. New York: Simon & Schuster.

Castaneda, Carlos (1977). *The Second Ring of Power*.

Castaneda, Carlos (1981). *The Eagle's Gift*.

Castaneda, Carlos (1984). *The Fire from Within*.

Castaneda, Carlos (1987). *The Power of Silence: Further Lessons with Don Juan*.

Castaneda, Carlos (1993). *The Art of Dreaming*.

Castaneda, Carlos (1998). *Magical Passes: The Practical Wisdom of the Shamans of Ancient Mexico*.

Castaneda, Carlos (1998). *The Wheel of Time: Shamans of Ancient Mexico. Their Thoughts about Life, Death and the Universe*.

Castaneda, Carlos (1998). *The Active Side of Infinity*.

6.2. Fuentes Secundarias Consultadas:

Abbasi, A. y Chen, H. (2005). Applying authorship analysis to extremist-group web forum messages. *IEEE Intelligent Systems*, 20(5).

Adair, D. (1944). "The Authorship of the Disputed Federalist Papers". *The William and Mary Quarterly*. Third Series, Vol.1, (Apr., 1944).

Adger, D. (2006). "Combinatory variation". *Journal of Linguistics*, 42.

Alcaraz Varó, E. y Martínez Linares, M. A. (1997). *Diccionario de lingüística moderna*. Barcelona: Editorial Ariel.

Applebome, P. (1998). "Carlos Castaneda, Mystical and Mysterious Writer, Dies". *The New York Times*. 20 de junio.

Applebome, P. (1998b). "Mystery Man's Death Can't End the Mystery; Fighting Over Carlos Castaneda's Legacy". *The New York Times*. 19 de Agosto.

Argamon, S. y Levitan, S. (2005). "Measuring the usefulness of function words for authorship attribution". *Proceedings of the Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing*.

Argamon, S.; Koppel, M.; Pennebaker, J. y Schler, J. (2009). "Automatically profiling the author of an anonymous text". *Communications of the ACM*.

Argamon, S. (2012). "Scalability Issues in Authorship Attribution. Kim Luycks". *Literary and Linguistic Computing*. Vol.27, nº1.

Atkinson, J. M.; Drew, P. (1979). *Order in court: the organization of interaction in judicial setting*. Londres: MacMillan.

Azuélos-Atias, S. (2007). *Legal Proofs of Criminal Intent*. Philadelphia: John Benjamin Publishing Company.

Baayen, H., van Halteran, H., Neijt, A. y Tweedie, F. (2002). "An Experiment in Authorship Attribution". *Journees internationales d'Analyse statistique des Donnees Textuelles* 6.

- Baayen, R., van Halteren, H. y Tweedie, F. (1996). "Outside the cave of shadows: Using syntactic annotation to enhance authorship attribution". *Literary and Linguistic Computing*, 11(3).
- Bailey, C.-J. (1973). *Variation and Linguistic Theory*. Arlington: Center for Applied Linguistics.
- Bailey, C. J. (1996). *Essays on Time-Based Linguistic Analysis*. Oxford: Oxford University Press.
- Baldick, C. (2001). *The Oxford Dictionary of Literary Terms*. Oxford: Oxford University Press.
- Bender, E. (2007). "Socially meaningful syntactic variation in sign-based grammar". *English Language and Linguistics* 11.
- Bengoechea, M. (2013). "Estudios de lingüística. Traducción e interpretación forense", en *Lingüística Forense. La Lingüística en el ámbito legal y policial*. Editores: Elena Garayzábal Heinze, Miriam Jiménez Bernal, Mercedes Reigosa Riveiros. Madrid: Euphonía Ediciones.
- Bennet, W. L. y Feldman, M. S. (1981). *Reconstructing reality in the courtroom*. Londres: Tavistock Publications.
- Bensalem I., Rosso P. y Chilhi S. (2014). "Intrinsic Plagiarism Detection using N-gram Classes." En: Proc. Int. Conf. on Empirical Methods in Natural Language Processing.
- Bergen, P. y Tiedemann, K. (2009). "The Drone War". *The New Republic*. June 3.
- Bergsma, S.; Post, M. y Yarowsky, D. (2012). "Stylometric Analysis of Scientific Articles". *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Bharati, A. (1980). "A Dual Mystical Fantasy". En *The Don Juan Papers. Further Castaneda Controversies*. Editor Richard de Mille. Lincoln, NE: iUniverse.
- Binongo, J. N. G. (2003). "Who wrote the 15th Book of Oz? An application of multivariate analysis to authorship attribution". *Chance*, 16(2).
- Blasco, J. y Ruiz Urbón, C. (2009). "Evaluación y cuantificación de algunas técnicas de "Atribución de autoría" en textos españoles". *Castilla. Estudios de Literatura*.

Bogdanova D., Rosso P. y Solorio T. (2014). “Exploring High-Level Features for Detecting Cyberpedophilia”. En: *Computer Speech & Language*, vol. 28, issue 1, pp. 108-120.

Borkman, D. (2011). LetterFrequency. Leipzig.

Bosmol Social Media News. (2013) “How big data analysis helped President Barack Obama defeat Romney in 2012 Elections”. Web.

Bradford, M.E. y Kirk, R. (1994). *Founding Fathers: Brief Lives of the Framers of the United States Constitution*. Lawrence: The University Press of Kansas.

Brainerd, B. (1973). “On the distinction between a novel and a romance: a discriminating analysis”. *Computing and the Humanities*. Nº 7.

Brainerd, B. (1974). *Weighing Evidence in Language and Literature: A Statistical Approach*. University of Toronto Press

Braz, A. (2011). “La lingüística forense i l’interrogatori policial a Espanya: Un enfocament lingüístic de les practiques actuals”. Llengua, Societat i Comunicació, nº 9. Web.

Brenan, M. (1997). “Of sorcery and dreams: An encounter with Carlos Castaneda”. *The Sun*. Septiembre.

Brennan, M.; Afroz, S. y Greenstadt, R. (2012). “Adversarial Stylometry: Circumventing Authorship Recognition to Preserve Privacy and Anonymity”. *ACM Transactions on Information and System Security*. Vol.15, nº3.

Brinegar, C. S. (1963). “Mark Twain and the Quintus Curtius Snodgrass Letters: A Statistical Test of Authorship”. *Journal of the American Statistical Association* 58.

Brown, P. y Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press

Buckleton, J.; Triggs, C. M. y Walsh, S. J. (2005) *Forensic DNA Evidence Interpretation*. Boca Raton; FL: CRC Press.

Burger, J. y Henderson, J. (2006). “An exploration of features for predicting blogger age”. *AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*.

- Burke, D. (2013). "The Cuckoo's Calling, by Robert Galbraith". *The Irish Times*. 27 de julio.
- Burrows, J. F. (1987). "Word patterns and story shapes: The statistical analysis of narrative style". *Literary and Linguistic Computing*, 2.
- Burrows, J. F. (1992). "Computers and the study of literature". En C. Butler, ed. *Computers and Written Text, Applied Language Studies*. Oxford: Blackwell.
- Burrows, J. F. (2002). "'Delta': A measure of stylistic difference and a guide to likely authorship". *Literary and Linguistic Computing*, 17(3).
- Burrows, J. F. y Craig, H. (2001). "Lucy Hutchinson and the Authorship of Two Seventeenth-Century Poems: A computational approach". *Eighteenth Century Studies*. Nº 16.
- Burrows, J. F. y Hassall, A. J. (1988). "Anna Boleyn and the Autenticity in Fielding's Feminine Narrative". *Eighteenth Century Studies*. 21.
- Bury, L. (2013). "Cuckoo's Calling by JK Rowling: did you know". *The Guardian*. 15 de julio.
- Carey, J. (1989). *Communication as Culture*. New York: Routledge
- Carmody, D. (2008). "Clay Felker, Magazine Pioner, Dies at 82" en *New York Times*. 1 de Julio.
- Carreter, L. (2008). *Diccionario de términos filológicos*. Madrid: Gredos.
- Chandler, D. (1994) "The Transmission Model of Communication". Aberystwith University. UK. Web.
- Chaski, C. E. (2005). "Who's at the Keyboard? Authorship Attribution in Digital Evidence Investigations". *International Journal of Digital Evidence*. Spring, Vol.4, Issue 5.
- Chaski, C. (2007). "Multilingual Forensic Author Identification through N-Gram Analysis". Presentado en 8th *Biennial Conference on Forensic Linguistics/Language and Law*. Seattle, WA.
- Chickering, L. A. (1988). "The Shaman of Ruthlessness". *Los Angeles Times*. 17 de enero.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1986). *Knowledge of Language: its Nature, Origin and Use*. New York: Praeger.

Chung, C.K.; Pennebaker, J.W. ([2003] 2007). "The psychological function of function words", en *Social Communication: Frontiers of Social Psychology*. New York: Psychology Press.

Cicres, J. y Turell, M. T. (2012). "Investigación, docencia y práctica profesional en lingüística forense", en Garayzábal Heinze, E.; Jiménez Bernal, M. y Reigosa Riveiros, M. Eds. *La lingüística forense: La lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.

Cleargreen (2013). Web.

Clement, R., y Sharp, D. (2003). "Ngram and Bayesian Classification of Documents." *Literary and Linguistic Computing*.

Cohn, N. (1976). "Tribal Rites of the New Saturday Night". *New York*. 7 de Junio.

Cole, S. A. (2002). *Suspect Identities. A History of Fingerprinting and Criminal Identification*. Boston: Harvard University Press.

Coleridge, S. T. ([1817] 2012). *Biographia Literaria*. Proyecto Gutenberg. Web.

Collins, S. (2008). *The Hunger Games*. New York: Scholastic Press.

Cook, B. (1973). "Is Carlos Castaneda for Real?". *National Observer*. 24 de febrero.

Coseriu, E. (1973). *Teoría del lenguaje y lingüística general*. Madrid: Editorial Gredos.

Coulthard, M. (2004). "Author identification, idiolect, and linguistic uniqueness". *Applied Linguistics*, 25(4).

Coulthard, M. (2005a). "The linguist as expert witness". *Linguistics and the Human Sciences*, 1(1).

Coulthard, Malcom (2005b). "Some forensic applications of descriptive linguistics". *Veredas*. Vol.9, nº1.

Coulthard, M. (2007). "Idiolect and uniqueness of encoding". En Coulthard, M. y Johnson, A. (2007). *An Introduction to Forensic Linguistics: Language in evidence*. Oxford: Routledge.

Coulthard, M. y Alison Johnson (2012). *The Routledge Handbook of Forensic Linguistics*. New York: Routledge.

Coyotl-Morales, R.M., Villaseñor-Pineda, L., Montes-y-Gómez, M., y Rosso, P. (2006). "Authorship attribution using word sequences". En *Proceedings of the 11th Iberoamerican Congress on Pattern Recognition*. Springer.

Craig, H. (1999). "Authorial attribution and computational stylistics: if you can tell authors apart, have you learned anything about them?" *Literary and Linguist Computing* 14.

Dale, E. y Chall, J. (1948). "A formula for predicting readability". *Educational Research Bulletin*, 27.

De Mille, R. (1976). *Castaneda's Journey. The Power and the Allegory*. Lincoln, NE: iUniverse.

De Mille, R. (1980). Editor. *The Don Juan Papers. Further Castaneda Controversies*. Lincoln, NE: iUniverse.

de Vel, O., Anderson, A., Corney y M., Mohay, G. M. (2001), "Mining e-mail content for author identification forensics". *SIGMOD Record* 30(4).

Delgado Romero, C. (2001). "La identificación de locutores en el ámbito Forense". Universidad Complutense Madrid. Tesis doctoral no publicada.

Delgado Romero, C. (2005). "Comentarios sobre el contexto actual de la Identificación Forense de Locutores". Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada. Barcelona.

Delgado Romero, C.; Márquez, M^a. C.; Olivas, M^a. C.; Barrios, L. (2009). "Identificación Forense de Locutores: Categorización de Parámetros Acústicos y Fono-articulatorios del español". *Revista de Lingüística Española*, Vol. 39-1.

Delgado Romero, C. (2012). "La pericia de identificación del habla: El papel fundamental del experto". En Garayzábal Heinze, E.; Jiménez Bernal, M. y Reigosa Riveiros, M. Eds. (2012). *La lingüística forense: La lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.

Derrida, J. ([1967] 1997). *Of Grammatology*. Baltimore: John Hopkins University Press.

De Saussure, F. ([1916] 2008). *Curso de Lingüística General*. Buenos Aires: Losada

Deutscher, G. (2010). *Through the Language Glass: Why the World Looks Different in Other Languages*. New York: Henry Hold and Company, LLC.

Eagleson, R. (1994). 'Forensic analysis of personal written texts: a case study', John Gibbons (ed.), *Language and the Law*, London: Longman,

Ebbinghaus, H. ([1885] 2013). "Introduction to *Memory* Hermann Ebbinghaus (1885/2013).

Eliade, M. (1964). *Shamanism: Archaic Techniques of Ecstasy*. New York: Pantheon Books.

Ellegard, A. (1962). "A Statistical Method for determining authorship: the 'Junius Letters' 1769-1772". En *Gothenburg Studies in English*. Universidad de Gothenburg.

Evans, S. P. y Skinner, K. (2005). *Jack The Ripper. Letters from Hell*. Londres: Sutton Pub Limited.

Executive Office of the President (2012). "Big Data Across the Federal Government". White House. September 26.

Fikes, J. C. (1993). *Carlos Castaneda. Academic Opportunism and the Psichedelic Sixties*. Victoria, B.C.: Millenia Press.

Finch, C. (2013). "The master is back in 'The Cuckoo's Calling'" *USA Today*. 16 de julio.

Fletcher, W. H. (2012). kfNgram.

Foster, D. W. (2001). *Author Unknown: On the Trail of Anonymous*. London: Macmillan.

Fucks, W. (1952). "On the mathematical analysis of Style". *Biometrika* 39.

Fucks, W. y Lauter, J. (1965). "Matematische analyse des literarschen stills". *Mathematik und Dichtung*. H. Kreuzer y R. Gunzenhäuser, eds. Munich.

Furst, P. (1972). *Flesh of the Gods. The Ritual Use of the Hallucinogens*. Westport: Praeger.

Garayzábal Heinze, E.; Jiménez Bernal, M. y Reigosa Riveiros, M. Eds. (2012). *La lingüística forense: La lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

- Garrett, M.F. (1982). "Production of Speech: Observations from Normal and Pathological Language Use", en A. Ellis (ed.) *Normality and Pathology in Cognitive Functions*. Londres: Academic Press.
- Gelb, I. (1952). *A Study of Writing*. Chicago: University of Chicago Press
- Gibbons, J. and Turell, M.T. Eds. (2008). *Dimensions of Forensic Linguistics*. Philadelphia: John Benjamin Publishing Company.
- Gil, J. y San Segundo, E. (2013). "La cualidad de voz en fonética judicial", en *Lingüística Forense. La Lingüística en el ámbito legal y policial*. Editores: Elena Garayzábal Heinze, Miriam Jiménez Bernal, Mercedes Reigosa Riveiros. Madrid: Euphonía Ediciones.
- Gladdis, K. (2013). "JK Rowling and the real-life whodunnit: Source of leak which revealed author wrote crime novel under male pseudonym is... solicitor from her OWN law firm". *Daily Mail*. 19 de julio.
- Gómez Jáuregui, V. (2007). *Tensegridad. Estructuras Tensegríticas en Ciencia y Arte*. Santander. Servicio de Publicaciones de la Universidad de Cantabria.
- Goldberg, Y. y Orwant, J. (2013). "A dataset of syntactic-Ngrams over time from very large corpus of English books". Web.
- Goldschmidt, W. ([1968] 1998), "Foreword". En Carlos Castaneda, *The Teachings of Don Juan: A Yaqui Way of Knowledge*, 30th Anniversary Edition. New York: Washington Square Press.
- Goodwin, W.; Linacre, A. y Hadi, S. (2007). *An Introduction to Forensic Genetics*. Hoboken, NJ: John Wiley & Sons.
- Grant, T. D. (2007). Quantifying evidence for forensic authorship analysis. *International Journal of Speech Language and the Law*, 14(1).
- Grant, T. (2008). "Approaching questions in forensic authorship analysis", en *Dimensions of Forensic Linguistics*. Eds. John Gibbons y Maria Teresa Turell. London: John Benjamins Publishing Company.
- Grieve, J. (2007). "Quantitative Authorship Attribution: An Evaluation of Techniques". *Literary and Linguistic Computing* 22 (3).
- Gumperz, J. ([1968] 2009). "The Speech Community", en Duranti, Alessandro (ed.) *Linguistic Anthropology: A reader* pg. 66-73. Chichester, West Sussex: John Wiley & Sons.

Gunning, R. (1952). *The Technique of Clear Writing*. New York: McGraw-Hill.

Gunning, R. (2006). "Clear Writing: How to achieve and measure readability". *The Writing Clinic*. November 2.

Hall, P. (2008). "Policiespeak", en Gibbons, J. and Turell, M.T. Eds. (2008). *Dimensions of Forensic Linguistics*. Philadelphia: John Benjamin Publishing Company.

Halliday, M.A.K. (1985). *An introduction to Functional grammar*. London: Edward Arnold.

Hardy, G. R.; Parsons, R. E. (2003). *Book of Mormon Plates and Record*. Web.

Harner, M. (1973). *Hallucinogens and Shamanism*. New York: Oxford University Press.

Harner, M. (1978). (Carta). *Los Angeles Times Book Review*. 1 de mayo.

Hilton, J. ([1982] 1997). "On Verifying Wordprint Studies", en en Noel B. Reynolds *Book of Mormon Authorship: New Light on Ancient Origins*. Provo, UT: Foundation for Ancient Research.

Hirst, G. y Feiguina, O. (2007). "Authorship attribution for small texts: literary and forensic experiments". SIGIR '07 Amsterdam. *Workshop on Plagiarism Analysis, Authorship Identification, and Near-Duplicate Detection*.

Hjelmslev, L. (1974). *Prolegómenos a una teoría del lenguaje*. 2ª ed., Madrid: Editorial Gredos.

Holland, N. N. (1975). *5 Readers Reading*. New Haven and London: Yale University Press.

Holmes, D. I. (1985). "The Analysis of Literary Style – A review". *Journal of the Royal Statistical Society. Series A (General)*, Vol.148, nº 4.

Holmes, D. I. (1991). "Vocabulary Richness and the Prophetic Voice". *Literary and Linguistic Computing*. Nº 6/4.

Holmes, D. I. (1998). "The Evolution of Stylometry in Humanities Scholarship". *Literary and Linguistic Computing* Vol.13, nº 3.

- Holmes, D. I., Gordon y L., Wilson, C. (2001). "A Widow and her Soldier: Stylometry and the American Civil War". *Literary and Linguistic Computing* 16(4).
- Honeybone, P. (2011). "Variation and linguistic theory", en Maguire, Warren & McMahon, April (eds.) *Analyxing Variation in English*. Cambridge: Cambridge University Press.
- Honore, A. (1979). "Some simple measures of richness of vocabulary". *Association for Literary and Linguistic Computing Bulletin*, 7(2).
- Hood, L. (2013). "Google's terms and conditions are less readable than Beowulf". *The Conversation*. 17 de octubre. Web.
- Hoover, D. (2004). Testing Burrows' Delta. *Literary and Linguistic Computing*, 19(4).
- Houvardas, J. y Stamatatos, E. (2006), "N-gram feature selection for authorship identification" En Proc. of the *12th Int. Conf. on Artificial Intelligence: Methodology, Systems, Applications*.
- Huber, R. A. y Headrick, A. M. (1999). *Handwriting Identification: Facts and Fundamentals*. Boca Ratón, FL: CRC Press.
- Hughes, R. (1973). "Don Juan and the Sorcerer's Apprentice". *Time*. 5 de marzo.
- Jackobson, R. (1960). "Closing statements: Linguistics and Poetics", en T.A. Sebeok *Style in language*. New York: Willey.
- Jeffries, L. y McIntyre, D. (2010). *Stylistics*. Cambridge: Cambridge University Press.
- Jiménez Bernal, M. (2012). "Estudios de lingüística, traducción e interpretación forenses", en Garayzábal Heinze, E.; Jiménez Bernal, M. y Reigosa Riveiros, M. Eds. (2012). *La lingüística forense: La lingüística en el ámbito legal y policial*. Madrid: Euphonia Ediciones.
- Jockers, M. L.; Witten, D. M. y Criddle, C. S. (2008). "Reassessing Authorship of the Book of Mormon Using Delta and Nearest Shrunken Centroid Classification". *Literary and Linguistic Computing*. 23/4.
- John, O. P. (1990), "The 'Big Five' factor taxonomy: Dimensions of personality in the natural language and in questionnaires", en Oliver P. John y Lawrence A. Pervin (eds.) *Handbook of Personality: Theory and Research*. New York: Guilford Press.

Joyce, G. H. (1919). *Principles of Logic*. London: Longman Green & Co. Web, 2013.

Joyce, J. (2008). "Bayes' Theorem". *The Stanford Encyclopedia of Philosophy* Edward N. Zalta (ed.), web.

Juola, P. (2008). "Authorship Attribution". *Information Retrieval*. Vol.1, nº3.

Juola, P. (2013a). "Rowling and "Galbraith": an authorial analysis". Web: *Language Log*. 16 de julio.

Juola, P., (2013b). "Stylometry and immigration: a case study". *Journal of Law and Policy*. Vol. XXI, nº2.

Juola, P. y Baayen, H. (2003). "A Controlled-corpus Experiment in Authorship Identification by Cross-Entropy". *Literary and Linguistic Computing* 20 (Suppl 1).

Kacmarcik, G. y Gamon, M. (2006). "Obfuscating Document Stylometry to Preserve Author Anonymity". *Association for Computational Linguistics*. Web.

Karlgren, J. and Cutting, D. (1994). "Recognizing text genres with simple metrics using discriminant analysis". En *Proceedings of the 15th Conference on Computational Linguistics*, Kyoto, Japan, August 1994, pp. 1071-1075.

Katritzky, L. (1996). *Johnson and the Letters of Junius; New Perspectives on an Old Enigma*. Nueva York: Peter Lang Publishing.

Kenny, A. (1978). *The Aristotelian Ethics: A Study of the Relationship between the Eudemian and Nicomachean Ethics of Aristotle*. Oxford: Clarendon Press.

Kenny, A. (1986). *A Stylometric Study of the New Testament*. New York: Oxford University Press.

Kerridge, J. (2013). "JK Rowling's crime novel: the verdict". *The Telegraph*. 10 de octubre.

Keselj, V.; Peng, F.; Cercone, N. y Thomas C. (2013). "N-gram-based author profiles for authorship attribution". *Pacific Association for Computational Linguistics*. Web.

Kessler, B.; Nunberg, G. y Schütze, H. (1997). "Automatic detection of genre". En *Proc. 35th Annual Meeting of the Association for Computational*

Linguistics and the 8th Meeting of the European Chapter of the Association for Computational Linguistics.

Kincaid, J.P.; Fishburne, R.P.; Rogers, R.L. y Chissom, B.S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula). *Navy Enlisted Personnel. Research Branch Report 8-75*. Chief of Naval Technical Training: Naval Air Station Memphis.

Kirk, P. L. (1953). *Crime investigation: physical evidence and the police laboratory*. New York: Interscience Publishers, Inc.

Kjell, B. (1994a). "Authorship attribution of text samples using neural networks and Bayesian classifiers". In IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX.

Kjell, B. (1994b). "Authorship determination using letter pair frequencies with neural network classifiers". *Literary and Linguistic Computing*, 9(2).

Kjell, B.; Woods, W. A. y Frieder, O. (1995), Information retrieval using letter tuples with neural network and nearest neighbor classifiers. En *IEEE International Conference on Systems, Man and Cybernetics*, volume 2. Vancouver, BC.

Kjetsaa, G. (1979). "And quite flows the Don' through the computer". *Ass. for Lit. and Linguist. Comput. Bull.*

Koppel, M.; Argamon, S. y Shimoni, A. (2002). "Automatically categorizing written texts by author gender". *Literary and Linguistic Computing*. Vol.17, nº4.

Koppel, M. y Schler, J. (2003). "Exploiting stylistic idiosyncrasies for authorship attribution". En *Proceedings of IJCAI'03 Workshop on Computational Approaches to Style Analysis and Synthesis*.

Koppel, M. y Schler, J. (2004). "Authorship Verification as a One-Class Classification Problem". ICML '04 Proceedings of the twenty-first international conference on Machine learning.

Koppel, M.; Schler, J.; Argamon, S. y Messeri, E. (2006). "Authorship attribution with thousands of candidate authors". En *Proceedings of the 29th ACM SIGIR*.

Koppel, M.; Schler, J. y Argamon, S. (2009). "Computational Methods in Authorship Attribution". *Journal of the American Society for Information and Technology*. Vol.60, nº1.

Koppel, M.; Schler, J. y Bonchek-Dokow, E. (2007). “Measuring differentiability: Unmasking candidate authors”. *Journal of Machine Learning Research*. N° 8.

La Barre, W. (1976). “Stinging criticism from the author of *The Peyote Cult*”. En *Seeing Castaneda*, Ed. D. Noel. New York: G. P. Putnam’s Sons.

Labov, W. ([1966] 2006). *The Social Stratification of English in New York City*. New York, NY: Cambridge University Press.

Labov, W. ([1972] 1991). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press

Lakoff, G. y Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: Chicago University Press.

Larsen, W. A.; Rencher, A. C. y Layton, T. ([1980] 1997). “Who Wrote the Book of Mormon? An Analysis of Wordprints”, en Noel B. Reynolds *Book of Mormon Authorship: New Light on Ancient Origins*. Provo, UT: Foundation for Ancient Research.

Ledger, G. y Merriam, T. (1994). “Shakespeare, Fletcher, and the Two Noble Kinsmen”. *Literature and Linguist Computing*.

Levs, J. (2013). “J.K. Rowling revealed as secret autor of crime novel”. Web: *CNN World*. 16 de julio.

Lewis, C. S. ([1961] 1992). *An Experiment in Criticism*. Cambridge, UK: Cambridge University Press.

Lin, Y.; Michel, J-B; Aiden, E. L.; Jon Orwant, J; Brockman, W. y Petrov, S. (2012). “Syntactic Annotations for the Google Books Ngram Corpus”. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 169–174, Jeju, Republic of Korea, July 2012.

Linn, W. A. (2011). *The Story of Mormons. From Date of their Origin to the Year 1901*. Public domain.

Lodge, K. (2009). *Fundamental Concepts in Phonology: Sameness and Difference*. Edinburgh: Edinburgh University Press.

Lombroso, C. (2009). *La escuela criminológica positivista*. Pamplona: Editorial Analecta.

Lopatka, M. y Peetz, M. H. (2011). “Biometric personal authentication using keystroke dynamics: A review”, en *Applied Soft Computing Journal*. Vol.11, Issue 2, Elsevier Science Publishers.

Love, H. (2002). *Attributing authorship. An introduction*. Cambridge: Cambridge University Press.

Lutoslawski, W. ([1897] 2006). *The Origin and Growth of Plato's Logic with an Account of Plato's Style and of the Chronology of his Writings*. Withefish, Montana: Kessinger Publishing LLC.

Luyckx, K. (2010). *Scalability Issues in Authorship Attribution*. Brussels: University Press Antwerp.

Luyckx, K. y Daelemans, W. (2005). "Shallow text analysis and machine learning for authorship attribution". En *Proceedings of the Fifteenth Meeting of Computational Linguistics in the Netherlands*.

Luyckx, K. y Daelemans, W. (2010). *The effect of author set size and data size in authorship attribution*". New York: Oxford University Press.

Marcus, G. y Fisher, M. (1986). *Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences*. Chicago: University of Chicago Press.

Marshall, R. (2007). "The Dark Legacy of Carlos Castaneda". *Salon*. 12 de abril.

Martindale, C. y McKenzie, D. (1995). "On the Utility of Content Analysis in Author Attribution: The 'Federalist'". *Computers and the Humanities*, 29, 259-270.

Martinet, A. ([1960] 1984). *Elementos de Lingüística General*. Madrid: Editorial Gredos.

Mathews, P. H. (1997). *The Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.

Mayzner, M. S. (1965). *Frequency Counts for Various Word-length and Letter-position Combinations*. Cambridge: Psychonomic Press.

McKeen, W. (1995). *Tom Wolfe*. New York, NY: Twaine Publishers.

McMenamin, G. R. (2002). *Forensic Linguistics*. New York: CRC Press

McMenamin, G. R. (2012). "Theory and Practice of Forensic Stylistics". En Malcom Coulthard y Alison Johnson (eds.). *The Routledge Handbook of Forensic Linguistics*. New York: Routledge.

Meyer, S. (2005). *Twilight*. New York: Little, Brown and Company.

- Mendelhall, T.C. (1887). "The characteristic curves of composition". *Science* 9, pp. 237-249.
- Merriam, T. y Matthews, R. (1994). "Neural computation in stylometry II: An application to the works of Shakespeare and Marlowe". *Literary and Linguistic Computing* 9, pp. 1-6.
- Michalos, A. (1969). *Principles of Logic*. Englewood Cliffs: Prentice-Hall.
- Midgley, L. C (1997) "Who really Wrote the Book of Mormon? The Critics and their Theories", en *Book of Mormon Authorship Revisited: The Evidence for Ancient Origins*, ed. Noel B. Reynolds. Provo, UT: Foundation for Ancient Research.
- Moehringer, J. R. (1998). "A Hushed Death for Mystic Author Carlos Castaneda". *Los Angeles Times*. 19 de junio.
- Monrose, F. y Rubin, A. D. (1999). "Keystroke dynamics as a biometric for authentication". *Future Generation Computer*. Elsevier.
- Morton, A. Q. (1965). "The Authorship of Greek Prose". *Journal of the Royal Statistical Society (A)*, 128.
- Mosteller, F. y Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalist*. Reading, Mass: Addison-Wesley.
- Murray, S. O. (1980). "The Invisibility of Scientific Scorn". En Richard de Mille ed. *The Don Juan Papers. Further Castaneda Controversies*. Lincoln, NE: iUniverse.
- Neon, Universidad. Freie Universität (2005). "TextStat". Holanda. Web.
- Nolan, F. (1994). «Auditory and acoustic analysis in speaker recognition». En J. Gibbons (Ed.) *Language and the law*. New York: Longman.
- Norvig, P. (2013). "English Letter Frequency Counts: Mayzner Revisited or ETAOIN SRHLDCU". Web.
- Oates, J. C. (1972). "Anthropology – or fiction?". (Carta). *New York Times Review*. 26 de noviembre.
- Oates, J. C. (1974). "Don Juan's Last Laugh". *Psychology Today*. 8 de septiembre.
- Olsson, J. (2008). *Forensic Linguistics: Second Edition*. New York: Continuum International Publishing Group.

Oppenheimer, W. (2013). "JK Rowling desvela que escribió con seudónimo una novela policiaca". *El País*. 14 de julio.

Ortega Dolz, P. (2013). "Big Brother is listening to you...". *El País*. 30 de abril.

Ortega Herraiz, J. M. (2011). "L'adaptació lingüística en els procediments judicials en què intervè un intérpret: realitat, motivació i conseqüències en el marc del procés penal a Espanya". *Llengua, Societat i Comunicació*, nº 9.

Patterson, W. P. (2008). *The Life & Teachings of Carlos Castaneda*. Fairfax, CA: Arete Communications Publishers.

Pavelec, D.; Oliveira, L.S.; Justino, E.; Nobre Neto, F.D.; Batista, L.V. (2009). "Compression and Stylometry for author identification". Proceedings of the 2009 International Joint Conference of Neural Networks.

Pawlowski, A. y Pacewicz, A. (2004). "Philosophe, helléniste ou fondateur sous-estimé de la stylométrie?" en Fögen, Thorstein and E.F.K. Koerner (eds.) *Historiographia Linguistica XXXI*: 2/3.

Peng, F.; Shuurmans, D.; y Wang, S. (2004). "Augmenting naive Bayes classifiers with statistical language models". *Information Retrieval Journal*, 7(1).

Pennebaker, J.W. y King, L.A. (1999). "Linguistic Styles: Language Use as an individual difference". *Journal of Personality and Social Psychology*. Vol.77, nº6.

Plimpton, G. (1966). "The Story Behind a Nonfiction Novel". *The New York Times*. 16 de Enero.

Publisher`s Weekly (2013). "The Cuckoo's Calling". Web, 25 de febrero.

Reno, S. (1980). "If Don Juan did not Exist, It Would Be Necessary to Invent Him". En Richard de Mille (ed.) *The Don Juan Papers. Further Castaneda Controversies*. Lincoln, NE: iUniverse.

Reyes A. y Rosso P. (2014). "On the Difficulty of Automatically Detecting Irony: Beyond a Simple Case of Negation." En: *Knowledge and Information Systems*, Sept., vol. 40, issue 3, pp. 595-614.

Rhodes, H. T. F. (1956). *Alphonse Bertillon: Father of Scientific Detection*. New York: Abelard-Schuman.

Riesman, P. (1972). "The collaboration of two men and a plant". *New York Times Book Review*. 22 de octubre.

Roszak, T. (1969). *The Making of a Counter Culture. Reflections on the Technocratic Society & Its Youthful Opposition*. Los Angeles: California University Press.

Rudman, J. (1998). "The state of authorship attribution studies: some problems and solutions". *Computer and the Humanities*. Nº 31.

Runyan, M. (1996). *A Magical Journey with Carlos Castaneda*. Victoria, Canada: Millenia Press.

Russell, B. ([1912] 2009). *The Problems of Philosophy*. Radford, VA: Wilder Publications.

Russell, B. (1952). "Is There a God?" Ensayo encargado, pero sin publicar para *Illustrated Magazine*. Web, 2013.

Sanderson, C. y Guenter, S. (2006). "Short Text Authorship Attribution via Sequence Kernels, Markov Chains and Author Unmasking: An Investigation", en *Int'l Conference on Empirical Methods in Natural Language Processing*.

Santana Lario, J. y Falces Sierra, M. (2012). "Any statement you make can be used against you in a court of law". *Introducción a la Lingüística Forense*. Universidad de Granada. Web.

Saussure, F. ([1916] 2008). *Curso de lingüística general*. Madrid: Alianza.

Savoy, J. (2012). "Authorship Attribution: A Comparative Study of Three Text Corpora and Three Languages". *Journal of Quantitative Linguistics*. Vol.19, nº2.

Schaalje, B. G.; Hilton, J. L. y Archer, J. B. (1997) "Comparative power of three author-attribution techniques for differentiating authors". *Journal of Book of Mormon Studies*. 6/1

Schaalje, B.; Roper, M. y Fields, P. J. (2012). "Stylometric Analyses of the Book of Mormon: A Short History". *Journal of the Book of Mormon and Other Restoration Scripture*. Vol. 21, nº1. Provo, UT: Maxwell Institute.

Schler, J.; Koppel, M.; Argamon, S. y Pennebaker, J. (2006). "Effects of age and gender on blogging." *AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*.

Scott, M. (1996). WordSmith 6.0. Oxford University Press

Shuy, R. W. (1984). Linguistics in other professions, *Ann. Rev. Anthropol.*

- Shuy, R. W. (1986). "Language and the law", *Ann. Rev. Appl. Linguis.*
- Shuy, R. W. (1993). *Language Crimes: the Use and Abuse of Language Evidence in the Courtroom*, Blackwell, Oxford.
- Shuy, R. W. (1998). *The Language of Confessions, Interrogation and Deception*, Sage Publications, Los Angeles.
- Sichel, H. S. (1974). "On a distribution representing sentence-length in written prose". *J. R. Statist. Soc. A*, 137.
- Sichel, H. S. (1986). "Word frequency distribution and type-token characteristics". *Mathematical Scientist*, 11.
- Simpson, P. (2009). *Stylistics. A Resource Book for Students*. New York: Routledge.
- Smith, E. H. (1983). "Recent experience and new developments of methods for the determination of authorship". *Literary and Linguistic Computing*. N° 11.
- Solam. L. M. (2013). "Intuition versus algorithm: The case of forensic authorship attribution". *Brooklin Journal of Law and Policy*. Vol. 21. 551. New York City.
- Spence, E. (2013). "The Real Winner of The Cuckoo's Calling Was Amazon, Not JK Rowling". *Forbes*. 15 de Julio.
- Stamatatos, E. (2009). "A survey of modern authorship attribution methods". *Journal of the American Society for Information Science and Technology*. Vol.60, nº3.
- Stamatatos, E., Fakotakis, N. y Kokkinakis, G. (2000). "Automatic text categorization in terms of genre and author". *Computational Linguistics*, 26(4), 471–495.
- Stamatatos, E., Fakotakis, N. y Kokkinakis, G. (2001). "Computer-based authorship attribution without lexical measures". *Computers and the Humanities*, 35(2), 193-214.
- Stanford Encyclopedia of Philosophy (2012). Web.
- Staniford, P. (1980). "I Come to Praise Carlos, Not to Bury don Juan". En *The Don Juan Papers. Further Castaneda Controversies*. Editor Richard de Mille. Lincoln, NE: iUniverse.

Stansall, B. (2013). “JK Rowling indemnizada por el bufete de abogados que reveló su seudónimo”. *El País*. 31 de julio.

Stefanova, M. (2009). “El potencial discriminatorio de las secuencias de categorías gramaticales en la atribución forense de autoría de textos en español”. Tesis doctoral. Barcelona: Universidad Pompeu Fabra.

Stygall, G. (1994). *Trial Language. Differential Discourse Processing and Discursive Formation*. Amsterdam and Philadelphia: John Benjamin.

Suzuki, D. T. (1994). *Manual of Zen Buddhism*. New York: Groove Press.

Tallentire, D. R. (1972). “An appraisal of methods and models in computational stylistics, with particular reference to author attribution”. *Ph.D. Thesis*, University of Cambridge.

Thompson, T. (1987). “The Evolution of Dandy Tom”. *Vanity Fair*. Octubre.

Turell, M. T. (2008). “La labor del lingüista forense en la determinación/atribución de autoría”. Barcelona: INGEFOR. ForensicLab.

Turell, M. T. (2011). “La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria de textos escrits”. *Llengua, Societat i Comunicació*, nº 9.

Turell M.T. y Rosso P. (2014). “Computational Approaches to Plagiarism Detection and Authorship Attribution in Real Forensic Cases.” En: *International Journal of Speech, Language and the Law*. Vol. 50. Pg. 21-28.

Tweedie, F. y Baayen, R. (1998). “How variable may a constant be? Measures of lexical richness in perspective”. *Computers and the Humanities*, 32(5).

Tweedie, F.; Holmes, D. y Corns, T. (1998). “The Provenance of *De Doctrina Christiana*, attributed to John Milton: A statistical investigation”. *Literary and Linguistic Computing*. 13.

Valenzuela, J. (1998). “El ‘brujo’ Carlos Castaneda muere con el misterio que caracterizó su vida y su obra”. *El País*. 20 de junio.

Vaux, B. (2008). “Why the phonological components must be serial and rule-based”. En Vaux, B. & Nevins, A. (eds.) *Rules, Constrains, and Phonological Phenomena*. Oxford: Oxford University Press.

Wake, W. C. (1957). “Sentence-length distributions of Greek authors”. *J. R. Statist. Soc. A*, 120, 331-346.

- Wales, K. (1989). *A Dictionary of Stylistics*. London: Longman.
- Wallace, A. (2003). *Sorcerer's Apprentice*. Berkeley: Frog Ltd.
- Waller-Davies, B. (2013). "Russells partner agrees to pay damages to JK Rowling over confidentiality breach". *The Lawyer*. 31 de julio.
- Wasson, R. G. (1972). "Review of *A Separate Reality*". *Economic Botany*. Enero-marzo.
- Wasson, R. G. (1974). "Review of *Tales of Power*". *Economic Botany*. Julio-septiembre.
- Watts, A. ([1957] 1999). *The Way of Zen*. New York: Pantheon Books.
- Waugh, S.; Adams, A. y Tweedie, F. J. (2000). "Computational stylistics using artificial Neural Networks". *Literary and Linguistic Computing*, 15(2) pp. 187-198.
- Weinrich, U.; Labov, W. y Herzog, M. I. ([1968] 2013). "Empirical foundations for a theory of Language Change". *Directions for Historical Linguistics: A Symposium*. Columbia University. Web.
- Williams, C. B. (1940). "A note on the statistical analysis of sentence-length as a criterion of literary style". *Biometrika*, 31.
- Wimsatt, W.K y M. Beardsley (1949). "The affective fallacy". *Sewanee Review*, vol. 57, nº 1.
- Wolfe, T. (1968). *The Pump-House Gang*. New York, NY: Farrar, Straus & Giroux.
- Wolfe, T. (1973). *The New Journalism*. New York, NY: Harper and Row.
- Yule, G. U. (1938). "On sentence-length as a statistical characteristic of style in prose, with application to two cases of disputed authorship". *Biometrika*, 30.
- Yule, G.U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.
- Zigdon, K. (2005). "Automatically determining an author's native language". M.Sc. Thesis, Dept. of Computer Science, Bar-Ilan University.

